

# **Predicción de la variación del Tipo de Cambio con Redes Neuronales: Rolling versus Recursividad.**

Seminario Para Optar al Título de Ingeniero Comercial

**Profesor:** Antonino Parisi F. Ph.D.

:

**Ignacio González R.**  
**José Mario Jiménez B.**  
**Santiago, Julio de 2003**



<b>Agradecimientos: . .</b>	<b>1</b>
<b>I. INTRODUCCIÓN . .</b>	<b>3</b>
<b>II. OBJETIVOS . .</b>	<b>5</b>
<b>A. OBJETIVOS GENERALES .</b>	<b>5</b>
<b>B. OBJETIVOS ESPECÍFICOS .</b>	<b>5</b>
<b>III. MARCO TEÓRICO .</b>	<b>7</b>
<b>A. TIPO DE CAMBIO: ANTECEDENTES GENERALES . .</b>	<b>7</b>
1. La economía chilena .	7
2. Condiciones de paridad entre monedas .	10
3. Tipo de cambio real y nominal . .	13
4. Regímenes cambiarios .	14
<b>B. REDES NEURONALES . .</b>	<b>15</b>
1. Las Neuronas . .	18
2. Funciones de Activación .	19
3. Características de la Redes Neuronales Artificiales (R.N.A.) .	24
4. Clasificación de las Redes Neuronales . .	25
5. Tipos de funcionamiento de las Redes Neuronales Artificiales .	26
6. Tipos de Redes Neuronales Artificiales .	30
7. Ventajas de las Redes Neuronales Artificiales . .	41
8. Desventajas de las Redes Neuronales .	42
<b>IV. MODELOS DE PREDICCIÓN DE REDES NEURONALES . .</b>	<b>45</b>
<b>A. METODOLOGÍA .</b>	<b>46</b>
1. Variables de entrada .	46
2. Modelos .	49
3. Rolling . .	50
4. Recursividad .	51
5. Direccional Accuracy Test (DA) . .	51

<b>B. ANÁLISIS DE RESULTADOS . .</b>	<b>52</b>
<b>V. CONCLUSIONES .</b>	<b>57</b>
<b>VI. BIBLIOGRAFÍA .</b>	<b>59</b>

## **Agradecimientos:**

*A nuestras familias por su cariño y apoyo constante.*

*Agradecemos al Ph.D. Antonino Parisi F. por su guía y amistad al dirigir nuestra tesis y por darnos una visión optimista de las finanzas.*

*A Pamela Gutiérrez V. e Isabel Soubllette por su compañía y apoyo incondicional.*

*A David Díaz y Edinson Cornejo por su generosa colaboración y disposición.*



# I. INTRODUCCIÓN

En las diferentes disciplinas de la ciencia, podemos ver que todos los días se hacen esfuerzos importantes para poder determinar con mayor certeza los fenómenos a los que nos vemos afectados. Para ello, se han implementado diversas técnicas de predicción para obtener resultados frente a estos nuevos eventos. Estos esfuerzos responden a la necesidad de las personas de disminuir el riesgo en la toma de decisiones y su aversión al riesgo en cuanto a las opciones que tienen que tomar.

En las finanzas la historia es muy parecida, durante mucho tiempo las personas han buscado maneras para poder acceder a mayor información que les permita poder tomar decisiones de una forma correcta, en donde las posibilidades de "equivocarse" sean las mínimas y en donde el poder tener éxito en la toma de decisiones sea lo más alto posible.

A medida que ha pasado el tiempo nos hemos visto expuestos a diversas técnicas para poder predecir los fenómenos futuros, ellas basadas en la premisa de que los elementos que suceden en la práctica, no son un efecto aleatorio, sino que representan de alguna manera tendencias que podrían ser explicadas de cierta forma por algún modelo.

Es así como nacen, por ejemplo, las técnicas con esquema y comportamiento lineal, dentro de las cuales podemos encontrar diversas técnicas que han ayudado a muchos inversionistas a lo largo de los últimos años.

Recientemente, se ha dado paso al modelamiento con técnicas no lineales para la predicción de mercados. Entre estas técnicas, podemos destacar las Redes Neuronales,

las cuales han sido toda una revolución en este sentido, ya que la manera en que ellas operan, es tratando de imitar el funcionamiento del cerebro humano, integrando la información pasada, "aprendiendo" de ella para luego tomar decisiones más certeras.

A pesar de todo, hemos visto que esta técnica arroja resultados poco certeros si es que no se seleccionan de manera correcta las variables de entrada a incluir, como así también la cantidad de datos que se van a utilizar en cada una de las diferentes etapas del funcionamiento del programa y su arquitectura. Es decir, esta nueva técnica no va a arrojar resultados correctos sólo por introducir los datos, sino que el poder identificar cuáles son las variables importantes que están influyendo en la variable de salida, la elección del conjunto de entrenamiento y su tamaño, su arquitectura, la elección de los parámetros de aprendizaje y el momento en que debemos detener el aprendizaje van a determinar el qué tan bien o mal va a funcionar el modelo.

En otras palabras, el planteamiento del problema va a ser un elemento clave para lograr el éxito al intentar predecir con mayor exactitud.

Además, hemos visto que la forma en que el programa estudia esta información, es también un factor muy importante para determinar con mayor certeza la predicción requerida. Es por esto que se han implementado técnicas en donde el programa va agregando datos a medida que va transcurriendo el tiempo, dándole un carácter más real, dado que las personas van haciendo lo mismo cuando toman decisiones, van agregando información actual para que su opción tenga mayor validez.

La necesidad de definir qué técnica es la más adecuada para utilizar en el modelo, es parte del motivo de nuestro estudio. Las técnicas que estudiaremos son las de funcionamiento Recursivo y Rolling. Ambas técnicas van agregando datos recientes e incorporándolos en el estudio de la nueva predicción, la diferencia va a radicar en la cantidad de datos que el programa va a ocupar para predecir la variable de salida.

Para el caso de la técnica Recursiva, el software va a ir agregando los nuevos datos, uno a uno, obteniendo cada vez un dato más que analizar con respecto a los casos anteriores. En el caso de la técnica Rolling, se van a ir agregando datos uno a uno, pero esta vez manteniendo constante el número de datos del conjunto de entrenamiento.

Nuestro estudio comparativo entre el funcionamiento Recursivo y Rolling lo vamos a realizar prediciendo el cambio de signo en el precio del tipo de cambio (US\$/Chilenos), viendo aquí un aporte importante a mucha gente que encuentra en la divisa un factor determinante para realizar negocios, como también en ser un aporte a los inversionistas que buscan un mayor grado de certidumbre a la hora de enfrentarse al valor del tipo de cambio.

## II. OBJETIVOS

### A. OBJETIVOS GENERALES

1. Crear modelos de predicción de la variación del tipo de cambio para el mercado chileno que logran un porcentaje de predicción de signo importante para la toma de decisiones.
2. Realizar un estudio comparativo de la predicción con Redes Neuronales utilizando un funcionamiento Recursivo versus Rolling.

### B. OBJETIVOS ESPECÍFICOS

1. Diseñar una base de datos con los valores del Tipo de Cambio y distintas variables financieras que influyen en los cambios del valor de la divisa.
2. Crear distintos modelos de predicción de la variación del tipo de cambio con Redes Neuronales Dinámicas con una arquitectura, variables y parámetros de aprendizaje adecuados.

3. Determinar cuál de los distintos modelos de redes neuronales logra un porcentaje de predicción de signo mayor.
4. Comparar que funcionamiento entrega mejores resultados al utilizarse en redes neuronales Rolling o Recursivo.
5. Encontrar el tamaño de Rolling que entregue el rendimiento más alto.
6. Evaluar si los resultados arrojados por los distintos modelos creados, son o no estadísticamente significativos.

## III. MARCO TEÓRICO

### A. TIPO DE CAMBIO: ANTECEDENTES GENERALES

#### 1. La economía chilena

---

La experiencia chilena en materia cambiaria ha sido muy variada. Podemos ver que han existido fluctuaciones significativas del tipo de cambio real durante más de 35 años, y cómo se ha pasado de períodos de apreciación a períodos de depreciación pronunciada.

Cabe recordar además que las dos últimas oportunidades en que se ha adoptado un esquema de tipo de cambio fijo, a principios de los sesenta y a fines de los setenta, la economía ha sufrido crisis de consideración.

La recuperación de la economía a la crisis de los ochenta estuvo basada en un fuerte crecimiento de las exportaciones y una notable depreciación del tipo de cambio real.

En los años más recientes, la acelerada depreciación de los 80' comenzó a revertirse en los 90'. Esto ha llevado, en el plano de la investigación, a la realización de importantes esfuerzos por analizar empíricamente los determinantes del tipo de cambio real. Así por ejemplo, se discute la responsabilidad de la política fiscal, los controles a la cuenta de capitales y los potenciales efectos de una modificación a la política arancelaria.

## **1.1 El contexto económico del mercado del dólar**

Las políticas económicas del gobierno tienen objetivos globales, entre otros, afianzar las condiciones de estabilidad macroeconómica y financiera del país (interna y externa en lo que respecta a sus relaciones con las economías en otros países) y fortalecer el conocimiento y disponibilidad de instrumentos para enfrentar las inevitables fluctuaciones de la economía tanto local como internacional, de manera de poder alcanzar los estándares económicos de países desarrollados.

Para alcanzar estos objetivos los agentes económicos deben solucionar, mejorar, controlar o ajustar las variables macroeconómicas más relevantes, que para simplificar las agrupamos en: crecimiento económico y productividad, inflación (IPC), desempleo, déficit en cuenta corriente, déficit público y finalmente el rol del Estado en la economía, y a su vez, plantear objetivos y establecer metas en torno a las variables mencionadas, las que se llevarán a cabo a través de políticas y medidas implementadas de manera firme, seria y consistente. La implementación de dichas políticas macroeconómicas está a cargo de dos de los principales agentes económicos del país, el Estado y el Banco Central de Chile.

Del Estado depende la política fiscal, tanto por el lado de los ingresos (principalmente tributos), como por el lado de los gastos. Sus objetivos son principalmente equilibrar y fortalecer las finanzas públicas; lograr tasas altas y estables de ahorro público (fuente de financiamiento de la inversión pública) y sustituir el endeudamiento externo. El instrumento utilizado para alcanzar estos objetivos es el manejo del gasto fiscal, el que se puede contraer o expandir según las contingencias económicas del momento.

Del Banco Central dependen la política monetaria, cambiaria y ciertos aspectos relacionados con el sistema financiero y del mercado de capitales (regulaciones a la cuenta de capitales). Estas políticas deben contribuir al logro de los objetivos que la ley le ha asignado al Banco Central: la estabilidad de la moneda, y por tanto, de los precios, esto es, alcanzar una inflación baja y estable, la preservación del sistema de pagos externos, lo que exige una posición externa sostenible, lo que se traduce entre otros elementos en mantener el déficit de cuenta corriente sobre una trayectoria sostenible; regular la expansión del gasto agregado interno y la preservación del sistema de pagos internos, lo que requiere de un sistema financiero sólido. Adicionalmente, la formulación de estas políticas afecta otras áreas de la economía como el crecimiento y el empleo.<sup>1 2</sup>

La conducción de la política monetaria en Chile se implementa, entre otros aspectos macroeconómicos, a través de la fijación de una meta de inflación anual. Este esquema permite un grado mayor de flexibilidad de las políticas con relación a los agregados monetarios o el tipo de cambio<sup>3</sup>. El otro instrumento a través del cual opera esta política es el control de la tasa de interés diaria, la que a su vez tiene efectos en las tasas de

<sup>1</sup> [Le Fort 97a]

<sup>2</sup> [Massad 98a]

<sup>3</sup> [Massad 98b]

interés reales de mediano y largo plazo, dejadas al control del mercado, que repercuten finalmente en el gasto interno privado (empresas y consumidores). En períodos de turbulencias el Banco Central mantiene un control estricto de la liquidez con el fin de no exacerbar estas volatilidades y ciclos especulativos contra el peso<sup>4</sup>. De esta manera la política monetaria se convierte en un aporte en el fortalecimiento de las bases de nuestra economía.

Otra de las herramientas sobre la cual el Banco Central tiene atribuciones es la cuenta de capitales. En este sentido las regulaciones que aplica el Banco Central tienden a evitar las oportunidades de arbitraje internacional de tasas de interés (interna y externa) y así permitir un aumento del tipo de cambio real. Las regulaciones aumentan el costo del financiamiento externo vía deuda de corto plazo, permitiendo a la autoridad mantener una tasa de interés interna más elevada sin que ella genere mayores entradas de capital, evitando una apreciación transitoria del tipo de cambio real. El fundamento macroeconómico de las regulaciones en la cuenta de capitales se encuentra en el hecho de que hace posible tasas de interés reales más altas que permitan un control más efectivo del gasto, mejorando la composición de los capitales externos al limitar su volatilidad. El registro de las transacciones de cuenta de capitales permite, entre otros aspectos, recoger información tanto estadística como para investigar operaciones de lavado de dinero y verificar los límites prudenciales en cuanto a activos externos que deben cumplir los intermediarios financieros. La efectividad de la regulación de la cuenta de capitales se observa en el diferencial de tasas de interés que existe, particularmente de corto plazo, entre Chile y los países desarrollados.

Las principales regulaciones se aplican a la entrada de capitales, incluyendo el encaje establecido para deuda e inversiones de cartera, el período de permanencia de un año para las inversiones en general, y las condiciones mínimas para emisiones de bonos y ADR. La salida de capital es, en general, completamente libre a través del mercado cambiario informal, aunque subsisten limitaciones al acceso de personas y empresas al mercado cambiario formal; por otra parte, bancos e inversionistas institucionales están afectos a montos máximos en sus tendencias de activos externos.<sup>5</sup>

La política cambiaria persigue también equilibrar las relaciones comerciales con el exterior y ayudar en la estabilización del ciclo económico. Esto se traduce en la determinación de los objetivos de esta política, tales como evitar que cambios bruscos en el tipo de cambio nominal saquen al tipo de cambio real de una trayectoria consistente con el equilibrio externo y una posición externa sostenible, y a su vez mantener controladas las expectativas cambiarias del sistema financiero. Esto último equivale a reducir la incertidumbre que existe en los mercados financieros internacionales con respecto a nuestra economía (estabilizar la moneda), que termina alterando las percepciones del riesgo cambiario, y por último, promover la integración financiera. El instrumento principal de esta política ha sido la definición de una banda cambiaria, la que orienta al mercado sobre los valores de la divisa que el Banco Central considera

---

<sup>4</sup> [Marshall 98]

<sup>5</sup> [Le Fort 97b]

compatibles con sus objetivos.

La importancia de conocer estas políticas y sus objetivos e instrumentos de operación reside en el hecho de que todas ellas manejan variables que están correlacionadas y, por lo tanto, cualquier modificación o alteración en una de ellas transmite sus efectos a las demás. De esta manera queda claro que todos los agentes económicos deben coordinar sus acciones dentro de un amplio contexto económico, con tal de producir efectos equilibrados y controlados.

## **2. Condiciones de paridad entre monedas**

---

En ausencia de intervención gubernamental, los tipos de cambio responden a las fuerzas de oferta y demanda, que a su vez dependen de las tasas de inflación relativas, de las tasas de interés y de las tasas de crecimiento del producto interno bruto.

En un mercado competitivo, con muchos compradores y vendedores con acceso a información, con activos y costos de transacción similares, debería darse ciertas condiciones de paridad que aseguren las condiciones de no arbitraje. De estos supuestos resultan una serie de relaciones teóricas que deberían cumplirse en un mercado en equilibrio. Estos modelos son más bien normativos, en la práctica son usados para explicar aisladamente los efectos de shocks económicos y tendencias de largo plazo basadas en factores estructurales. Las relaciones fundamentales son:

Paridad del Poder de Compra (PPC)

Efecto Fisher (EF)

Efecto Fisher Internacional (EFI)

Paridad de las Tasas de Interés (PTI)

Tasa Forward como predictor insesgado de la futura tasa Spot (EPIS)

### **2.1 Teoría de la Paridad del Poder de Compra**

Sostiene que una unidad monetaria cualquiera debiera tener el mismo poder de compra en todo el mundo. En otros términos, que los precios ajustados por el tipo de cambio debieran ser idénticos en todas partes, con ello los tipos de cambio se ajustarán para reflejar cambios en los niveles de precios entre los distintos países.

Formalmente, si  $i_h$  e  $i_f$  son las tasas de inflación local y de otra economía respectivamente, siendo  $e_0$  el precio de la divisa en la moneda local al principio y  $e_t$  el precio de la misma en  $t$  períodos, deberá darse la condición:

$$\frac{e_t}{e_0} = \frac{(1 + i_h)^t}{(1 + i_f)^t} \quad (1)$$

De ahí que las monedas con altas tasas de inflación deberán devaluarse con respecto a otras que presentan índices inflacionarios menores. De esta relación también

se desprende que el tipo de cambio real sea el tipo de cambio nominal ajustado por la inflación que afecta a ambas monedas para un período determinado. Esta teoría es utilizada por bancos centrales como una guía para establecer paridades cambiarias acordes a la situación macroeconómica de un país. También, a pesar de sus limitaciones, es utilizada para estimaciones de largo plazo. En la ecuación siguiente se muestra esta relación, donde  $e'_t$  es el tipo de cambio real,  $e_t$  es el tipo de cambio nominal y la notación en los otros términos es similar a la de la ecuación anterior.

$$e'_t = e_t \frac{(1 + i_h)^t}{(1 + i_f)^t} \quad (2)$$

Las limitaciones de este enfoque están fundamentalmente en que se asume cuentas de capitales abiertas en todo el mundo, desprecia los costos de transporte y de transacción y asume que todos los índices inflacionarios se calculan en base a una misma canasta de bienes. Estos factores hacen que los tipos de cambio difieran de esta paridad, en especial el supuesto de apertura de cuentas de capitales, ya que las barreras que imponen gran parte de los países introducen fuertes distorsiones a la ley de precio único.

En todo caso, ante fuertes procesos inflacionarios o economías pequeñas o abiertas, este criterio se comporta mejor y entrega adecuadas tendencias de largo plazo.

## 2.2 El efecto Fisher

Relaciona tasas de interés real y nominal a través de la ecuación:

$$(1 + r_n) = (1 + r_R)(1 + i) \quad (3)$$

En la fórmula anterior  $r_n$  es la tasa de interés nominal,  $r_R$  la tasa de interés real e  $i$  la tasa de inflación. Este es un criterio ampliamente conocido en finanzas, además la teoría macroeconómica explica que escenarios con altas tasas inflacionarias están vinculados a fuertes tasas de interés a raíz de la aceleración del gasto.

## 2.3 El efecto Fisher internacional

Es una síntesis de la paridad del poder de compra (PPC) y el efecto Fisher generalizado, es decir, un alza en la inflación vendrá acompañada de un aumento en las tasas de interés en el mercado local y una depreciación de la moneda, esta relación se ilustra en la ecuación que sigue a continuación.

$$\frac{(1 + r_h)^t}{(1 + r_f)^t} = \frac{e_t}{e_0} \quad (4)$$

Donde el primer término es el cociente entre los tipos de interés compuesto en  $t$  períodos para la moneda local y la extranjera; el numerador del segundo término es el tipo de cambio esperado para el período  $t$  y el denominador de éste es el tipo de cambio spot.

En resumen, este enfoque establece que monedas con bajos tipos de interés deberán apreciarse con respecto a aquellas que presentan mayores tasas de interés. Es un indicador ampliamente utilizado para el corto plazo, pero requiere distinguir claramente si los diferenciales entre tasas de interés nominal se dan debido a alzas en las expectativas inflacionarias o simplemente a variaciones en los tipos de interés reales.

## **2.4 Teoría de paridad en las tasas de interés**

Según esta teoría, la moneda de un país con una tasa de interés baja debería cotizarse con un premio sobre la moneda de un país que tenga una tasa de interés más alta. Así, en un mercado eficiente, sin costos de transacción, el diferencial de tasas de interés debería compensar las diferencias entre los tipos forward y spot:

$$\frac{1 + r_h}{1 + r_f} = \frac{f_1}{e_0} \quad (5)$$

La paridad de las tasas de interés es una de las relaciones mejor documentadas en las finanzas internacionales. Así en la práctica, la tasa forward se calcula a partir del diferencial de intereses entre dos monedas usando la condición de no arbitraje. Sin embargo, en la práctica suelen ocurrir desviaciones debido a restricciones en la cartera de capitales, restricciones a la inversión extranjera y costos de transacción. La normativa para el ingreso de capitales extranjeros a Chile es uno de estos ejemplos, donde el encaje busca mantener tasas de interés mayores sin afectar el tipo de cambio.

## **2.5 Relación entre tasa forward y tasa spot futura:**

En ausencia de la intervención gubernamental en el mercado, tanto las tasas spot como las tasas forward sufren una fuerte influencia por las expectativas de los eventos futuros y se mueven manteniendo un lazo basado en los diferenciales de los intereses. Cualquier información en cuanto al cambio de las tasas de interés se refleja inmediatamente en ambas. Sólo se logra un equilibrio cuando los diferenciales en los tipos de forward compensan las variaciones en el tipo de cambio spot, de lo contrario se tendrían condiciones de arbitraje, de hecho las presiones que afecten al mercado spot serán transmitidas al mercado forward y viceversa.

De acuerdo a lo anterior y a su relación con el efecto Fisher y paridad del poder de compra, la tasa forward debería reflejar la tasa spot esperada en la fecha establecida en el contrato forward.

$$f_t = \bar{e}_t \quad (6)$$

Lo anterior indica que en este mercado no debería haber premios por riesgo involucrados. Los estudios más recientes demuestran que puede haberlos, siendo a veces positivos y otros negativos. En la práctica si el motivo de las transacciones forward es la cobertura, los hedgers aceptarán un precio forward mayor que el valor esperado de la tasa spot futura; por otra parte, los productores exigirán una tasa forward menor al valor esperado de la tasa spot futura.

De acuerdo a ello, los premios deberían tender a hacerse pequeños en un mercado profundo para incentivar a productores y consumidores a intervenir. El hecho de que en general los mercados de divisas presentan mayor profundidad que otros mercados hace pensar que el efecto de estos premios no es significativo.<sup>6</sup>

### 3. Tipo de cambio real y nominal

El tipo de cambio nominal, asociado al mercado de las divisas, debe sus movimientos al propio mercado y a las expectativas de variaciones futuras influenciadas por factores políticos, económicos y sociales del momento. Así, los inversionistas prefieren obviamente tener monedas de bajo riesgo, las cuales están asociadas a países más estables política y económicamente. Por ejemplo, los problemas económicos internacionales generaron presiones de depreciación que sorprendieron al mercado. Esto llevó a importantes empresas nacionales a cubrir sus posiciones en moneda extranjera a través de operaciones *Forward* de compra de dólares, lo cual indirectamente presionó al alza del tipo de cambio spot, por la necesidad de cubrir riesgo cambiario asociado a los contratos.

El tipo de cambio real, asociado al mercado de bienes, es un indicador que nos aproxima al grado de competitividad internacional de los productos nacionales (sector exportador). Lo definimos como el cociente de los precios domésticos de los bienes transables dividido por los precios domésticos de los bienes no transables. Así una caída en el indicador, generado por un aumento del precio de los bienes no transables, indica que han aumentado domésticamente los costos de los productos transable, empeorando la competitividad del sector exportador chileno, con respecto al resto del mundo.<sup>7</sup>

La correlación entre tipo de cambio nominal y real se explica por las teorías de equilibrio y desequilibrio de los tipos de cambio. Esta última sostiene que por fricciones en la economía los precios de los bienes se ajustan más lentamente que los precios de las monedas. Esta diferencia en los ajustes de los distintos mercados indica que los movimientos del tipo de cambio nominal generan variaciones del tipo de cambio real en el corto plazo y que podrían exceder los equilibrios de corto plazo. Esto genera que los

<sup>6</sup> [Virreira 96a]

<sup>7</sup> [Sanhueza 98]

equilibrios en el mercado de las divisas afecten la competitividad interna, pero no alcanzan a modificar las ventajas comparativas. Por lo tanto, un movimiento efectivo del tipo de cambio real se producirá ante la modificación en los precios relativos y tipo de cambio nominal. Es decir, los movimientos del tipo de cambio nominal no alteran los precios relativos, pero son parte del proceso por medio del cual los cambios ocurren en equilibrio.<sup>8</sup>

Esto no significa que el Banco Central pueda determinar ni mantener el tipo de cambio real en un cierto nivel, ya que éste es el resultado del equilibrio en el mercado de bienes. El tipo de cambio real responde a factores tales como la productividad, la apertura de nuevos mercados, el desarrollo de nuevos productos, el nivel de los términos de intercambio y la brecha entre la evolución del ingreso y la evolución del gasto o la brecha entre ahorro e inversión. Por otra parte, el tipo de cambio nominal responde a la oferta y demanda de divisas de agentes económicos, tales como, importadores, exportadores, bancos comerciales, inversionistas entre otros, y el Banco Central que compra o vende divisas como una herramienta de intervención cambiaria.

## **4. Regímenes cambiarios**

---

Existen varios patrones relacionados al grado de movilidad del tipo de cambio y a las posibilidades de intervención del Banco Central en la fijación de su valor, definido como el precio relativo de la moneda local y alguna divisa. En nuestro caso, lo definimos como la cantidad de pesos que nos cuesta un dólar de Estados Unidos (\$/US\$), o cualquier otra moneda extranjera.

### **4.1 Tipo de cambio fijo**

Bajo este sistema la autoridad monetaria se compromete a mantener su moneda a una paridad fijada, por ejemplo \$700/US\$1. Si existen amenazas macroeconómicas que tiendan a sacar de ese valor al tipo de cambio, la autoridad podría decretar realizar revaluaciones o devaluaciones según sea el caso. Dentro de este régimen, también existe la posibilidad de una banda cambiaria en torno al valor fijado, permitiendo una flotación más flexible. Así, la intervención se limita a los casos en que el tipo de cambio amenace con salirse de la banda.<sup>9</sup>

### **4.2 Esquema de bandas cambiarias**

Según el trabajo de Krugman se demuestra que la mera existencia de límites a la fluctuación del tipo de cambio, las cuales la autoridad se compromete a defender vía intervenciones en el mercado cambiario, conlleva a importantes efectos en la dinámica del tipo de cambio.

Él parte del supuesto de que el tipo de cambio, como cualquier otro precio, depende

<sup>8</sup> [Virreira 96b]

<sup>9</sup> [O'Kean 94]

de ciertas variables fundamentales contemporáneas, así como de la expectativa de los valores futuros. A modo de simplificación, se asume que el tipo de cambio depende linealmente de un agregado “fundamental”, el cual incorpora los distintos determinantes fundamentales del tipo de cambio y de la variación esperada del tipo de cambio.

El Banco Central, a través del control de la oferta monetaria, puede controlar el “agregado fundamental”. En efecto, mediante una expansión monetaria puede provocar un incremento en el nivel del tipo de cambio, mientras que a través de una contracción puede provocar que el tipo de cambio caiga. Bajo la existencia de un esquema de bandas para el tipo de cambio, la autoridad monetaria interviene, expandiendo o contrayendo la oferta de dinero, de tal manera de mantener el tipo de cambio nominal dentro de un rango previamente anunciado.<sup>10</sup>

### 4.3 Tipo de cambio flexible

Si el tipo de cambio se fija en el mercado de las divisas libremente por la interacción de oferentes y demandantes, decimos que el tipo de cambio es flexible y no hay compromiso por parte del Banco Central de sostener un determinado tipo de cambio.

Si no hay ningún tipo de intervención, el esquema se define como flotación limpia. En caso contrario, cuando el Banco Central compra o vende divisas, el esquema es llamado flotación sucia<sup>11</sup>. Así, siendo la oferta de dinero constante, los cambios en el agregado fundamental dependen sólo de la velocidad de circulación, por lo que presenta un movimiento browniano sin tendencia. De modo equivalente, el tipo de cambio depende linealmente del agregado fundamental y de las expectativas futuras de cambios en el tipo de cambio. Si el tipo de cambio es un movimiento browniano sin tendencias, entonces las expectativas futuras de cambios en la paridad cambiarias son nulas, y el tipo de cambio dependería sólo del agregado fundamental de modo lineal.

Debido que el agregado fundamental es un movimiento browniano sin tendencia, esto es consistente con que el tipo de cambio sea un movimiento de este mismo tipo.<sup>12</sup>

## B. REDES NEURONALES

Las Redes Neuronales son un nuevo paradigma en procesamiento de datos con grandes perspectivas para manejar muchos de los más desafiantes problemas computacionales. Mientras las técnicas de inteligencia artificial han ayudado a expandir las capacidades de los computadores existentes automatizando los procesos de razonamiento humano a nivel de *software*, las redes neuronales son un intento de imitar las capacidades de

<sup>10</sup> [Magendzo 96a]

<sup>11</sup> [Virreira 96c]

<sup>12</sup> [Magendzo 96b]

procesamiento del cerebro a nivel de *hardware*; esto es, a nivel de las neuronas individuales y su comportamiento general cuando se combinan en grandes redes altamente interconectadas.

Las redes neuronales se han aplicado en numerosas áreas que involucran el procesamiento de señales, el reconocimiento de patrones complejos y las tareas de clasificación, así como el procesamiento de imágenes, el reconocimiento de voz y caracteres, filtrado de ruido, compresión de datos, predicción de series de tiempo y optimización. Otra importante área de investigación es el control automático, donde las redes neuronales han sido investigadas para el control adaptativo de robots y vehículos autónomos, y la identificación y control de procesos complejos o desconocidos. La capacidad de las redes neuronales para el aprendizaje adaptativo, mapeo de señales no lineales y tolerancia a fallas están siendo explotadas para llenar muchas de las persistentes fallas presentes en las técnicas clásicas de control.<sup>13</sup>

Según Martín del Brío y Sanz una definición de redes neuronales sería la siguiente:

“Las redes neuronales artificiales son sistemas de procesamiento que copian esquemáticamente la estructura neuronal del cerebro para tratar de reproducir sus capacidades, como por ejemplo, aprender de la experiencia a partir de las señales o datos del exterior, con el objetivo de construir sistemas de procesamiento de la información paralelos, distribuidos y adaptativos”.<sup>14</sup>

Las redes neuronales son sistemas de procesamiento que tratan de reproducir la capacidad que tiene el cerebro humano para reconocer patrones, hacer predicciones o tomar una decisión en base a la experiencia pasada. Así, mientras el cerebro humano confía en la estimulación neuronal, las redes neuronales actúan sobre un conjunto de datos para aprender. Dichas redes destacan por su desempeño en la solución de complejos problemas, donde el reconocimiento de modelos o comportamientos es importante.

Uno de los primeros modelos matemáticos de una neurona fue el propuesto por McCulloch y Pitts en 1943. En este modelo, ilustrado en la figura a continuación, cada neurona consta de un conjunto de entradas,  $S_i$ , y una sola salida  $S_j$ . Cada entrada  $i$  está afectada por un coeficiente que se denomina peso y se representa por la letra  $W_{ij}$ . El subíndice  $i$  refleja que el peso afecta a la entrada  $i$ , y el subíndice  $j$  que se trata de la neurona  $j$ .

---

<sup>13</sup> [Jiménez 96a]

<sup>14</sup> [Reyes 02a]

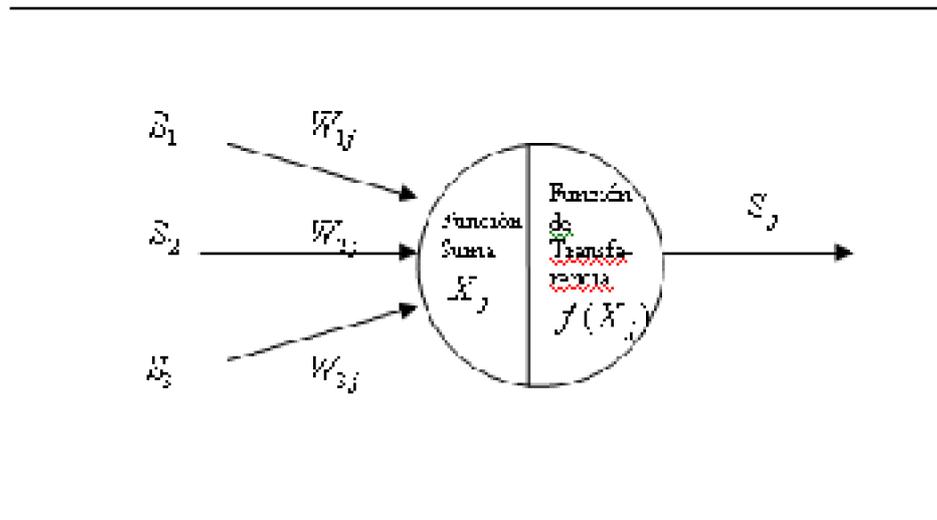


Figura 1: Modelo simple de una Neurona

La estructura clásica de una red neuronal se ve reflejada en la siguiente arquitectura:

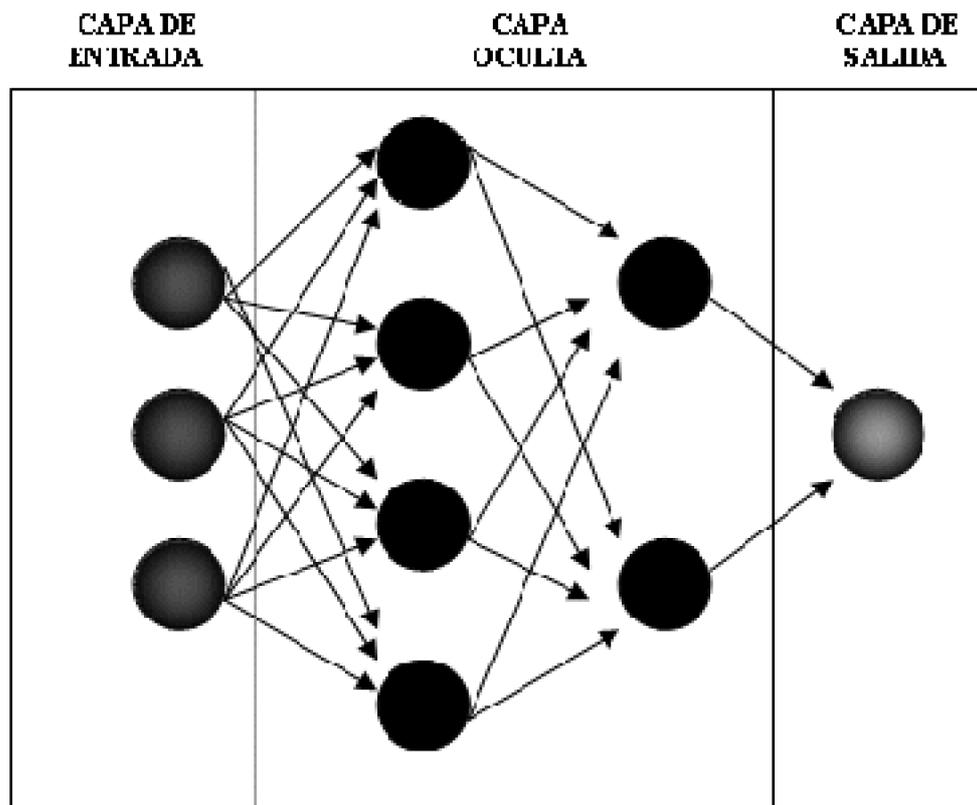


Figura 2: Estructura clásica de una red neuronal

Donde se pueden distinguir como parte de su diseño:

Capa de entrada:

Destinada a recibir la información para la resolución del problema (variables independientes). Tiene como características:

- a) Ser de carácter única.
- b) Número de neuronas definido por la cantidad de inputs.
- c) Propaga los datos a la capa siguiente sin proceso alguno, salvo la aplicación de una función escalar.

Capa Oculta:

Donde se lleva a efecto el procesamiento de los datos. Tiene como características:

- a) La no cualidad de única, pudiendo existir más de una en la arquitectura de red.
- b) Número de neuronas definido por quien aplica la red.
- c) Los datos se propagan a la capa siguiente previo acto de una función de activación.

Capa de Salida:

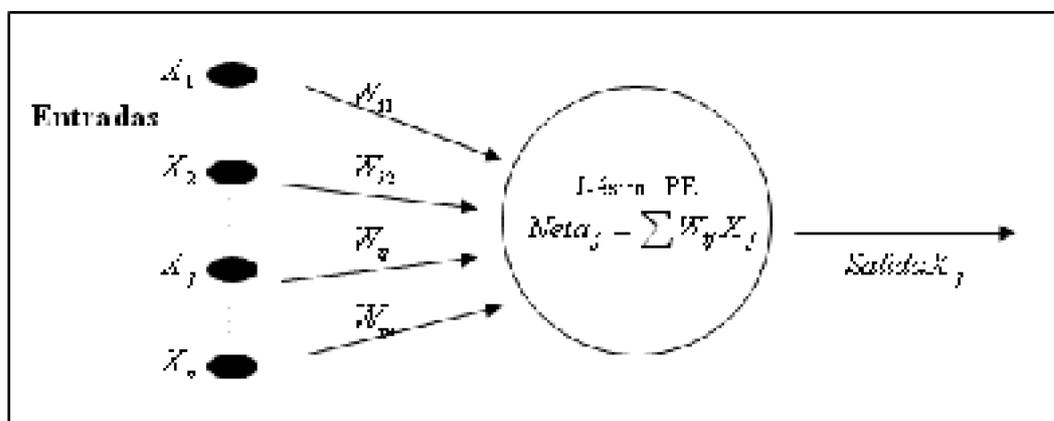
En las que se presenta el resultado entregado por la red neuronal. Tiene como características:

- a) Ser de carácter única.
- b) Número de neuronas definido por la cantidad de outputs.
- c) Neuronas aplican una función de activación a los datos que reciben, antes de dar el producto final.<sup>15</sup>

## 1. Las Neuronas

---

La neurona, también llamada procesador elemental (PE), es un dispositivo simple de cálculo que a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una única respuesta o salida (ver figura 3). Los procesadores elementales se agrupan formando capas, las cuales a su vez conforman las llamadas redes neuronales.



<sup>15</sup> [EspinozaMu 02a]

Figura 3: La Neurona

Al igual que una neurona biológica, el PE tiene muchas entradas pero una sola salida, que se puede aplicar a muchos otros PEs de la red. La salida del  $j$ -ésimo nodo es  $X_j$ , del mismo modo que la salida generada por el  $i$ -ésimo nodo se denomina  $X_i$ . Cada conexión con el  $i$ -ésimo PE tiene asociada a él un valor llamado *peso* o *intensidad de conexión*. El peso de la conexión procedente del  $j$ -ésimo nodo y que llega al  $i$ -ésimo nodo se denomina  $W_{ij}$ . Una conexión de entrada puede ser excitatoria (pesos positivos) o inhibitoria (pesos negativos). Cada PE determina un valor de entrada neto basándose en todas las conexiones de entrada. La forma más común es calcular el valor de entrada neto sumando los valores de entrada, ponderados mediante sus pesos correspondientes. Es decir, la entrada neta de la  $i$ -ésima unidad se puede escribir de la siguiente forma:

$$\text{Neta}_i = \sum_j X_j W_{ij} \quad (7)$$

En donde el índice  $j$  recorre todas las conexiones que posea el PE. Una vez que la entrada neta ha sido calculada, se transforma en el valor de activación para ese PE.

$$a_i(t) = F_i(a_i(t-1), \text{neta}_i(t)) \quad (8)$$

En la mayoría de los casos, la activación y la entrada neta son idénticas, y los términos suelen utilizarse de manera intercambiable.

Una vez que se ha calculado la activación del PE, se puede determinar el valor de salida aplicando la función de salida que puede escribirse como:

$$X_i = f_i(\text{neta}_i) \quad (9)$$

## 2. Funciones de Activación

Las funciones de activación son los elementos de la topología de la red que determinan en que nivel de datos se quiere poner el mayor énfasis. Por ejemplo, si se utiliza una función lineal todos los datos son tratados de la misma manera; por el contrario, si se usara una función tangente hiperbólica, ésta consideraría más relevante la información alrededor de cero que aquella que se encuentra en los extremos de uno y menos uno.

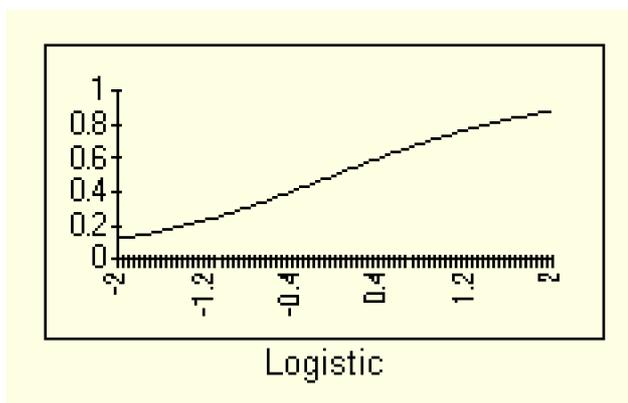
Los tipos de función de activación que podrían utilizarse son ilimitados. Sin embargo las más comunes (más utilizadas) son las siguientes:

### 2.1 Función Logística

La función Logística o Sigmoide es la más común en la mayor parte de las redes neuronales con aplicaciones financieras. Su rango es (0,1). Se utiliza para concentrar el

aprendizaje en valores no extremos en donde deberían encontrarse la mayor parte de los casos, por ello es de especial utilidad cuando los outputs son categorías. En términos matemáticos su expresión es:

$$F(x) = \frac{1}{(1 + e^{-x})} \quad (10)$$

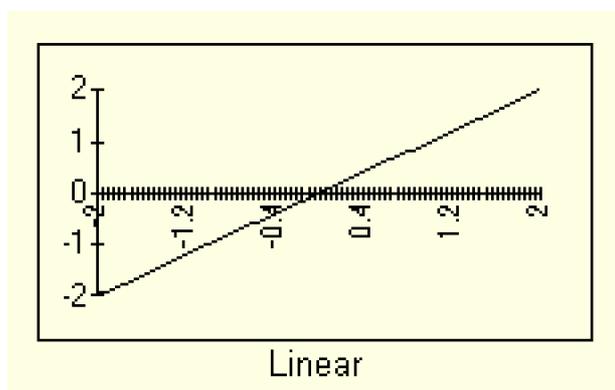


*Figura 4: Función Logística*

## 2.2 Función Lineal

La función Lineal es especialmente útil en aquellos casos en que el output es una variable continua o en aquellos en que se desea que la red aprenda los eventos menos frecuentes, es decir, aquellos que coincidirán con valores alejados a cero en la escala. Ello se debe a que a diferencia de la función Sigmoide, la Lineal no se hace menos sensible al alejarse de cero. La expresión matemática es:

$$F(x) = x \quad (11)$$



*Figura 5: Función Lineal*

### 2.3 Función Tangente Hiperbólica

La Función Tangente Hiperbólica tiene las mismas propiedades que la Logística, sin embargo, el rango de salida de esta función permite respuestas simétricas (-1,1); manteniendo una intermedia en cero. Esto puede ser de especial interés al escalar los datos en la primera capa de la red, pues en muchos casos puede necesitarse atenuar los valores exageradamente altos o bajos, pero diferenciándolos de un nivel donde la información sea irrelevante en donde se tendrá como valor cero.

Esta función suele converger antes que la función Logística, sin embargo, no necesariamente generaliza igualmente bien. En términos matemáticos la función es:

$$F(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (12)$$

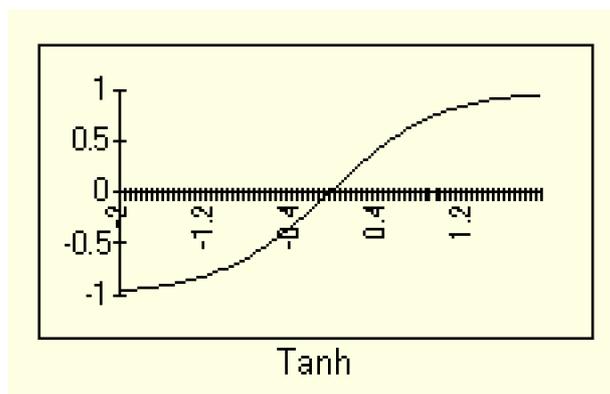


Figura 6: Función Tangente Hiperbólica

### 2.4 Tangente Hiperbólica 1,5

La Función Tangente Hiperbólica con parámetro 1,5, concentra más el aprendizaje en valores cercanos a cero amplificando el valor de la variable de entrada por 1,5. La diferencia con la Tangente Hiperbólica convencional debe probarse empíricamente. La fórmula es:

$$F(x) = \frac{(e^{1.5x} - e^{-1.5x})}{(e^{1.5x} + e^{-1.5x})} \quad (13)$$

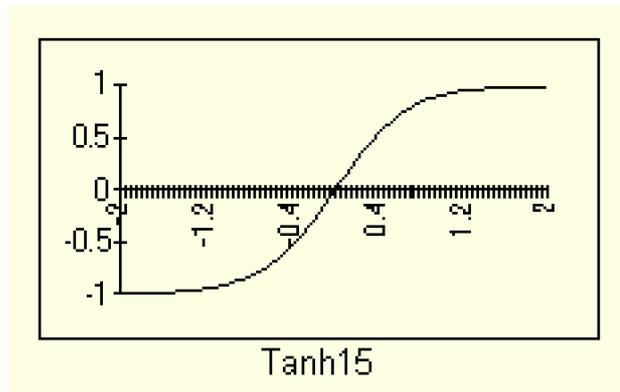


Figura 7: Función Tangente Hiperbólica 1,5

## 2.5 Seno

El uso de la función Seno es sólo una alternativa adicional a probar, su forma en algunos casos puede implicar la obtención de un mejor ajuste. En el gráfico se observa esta función. Su fórmula es:

$$F(x) = \text{Sen}(x) \quad (14)$$

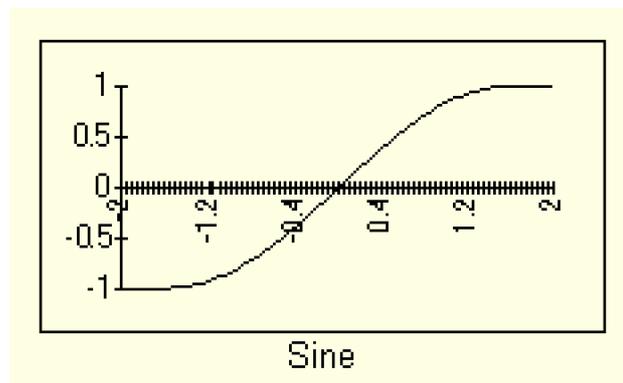


Figura 8: Función Seno

## 2.6 Logística Simétrica

Esta función es similar a la Sigmoide, pero se mueve en el rango (-1,1). La aplicación de esta función de transferencia debe considerarse de acuerdo al escalamiento de los datos. Su expresión matemática es la siguiente:

$$F(x) = \frac{2}{e^{-x}} - 1 \quad (15)$$

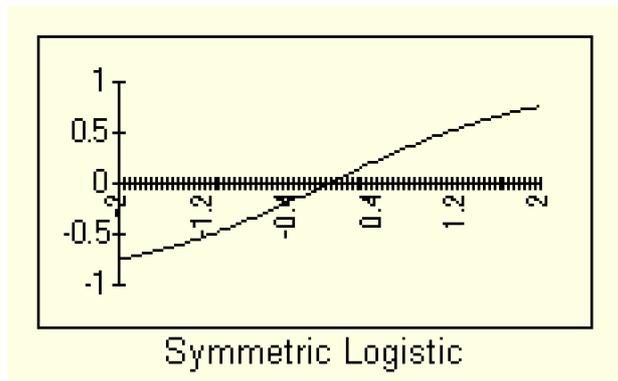


Figura 9: Función Logística Simétrica

## 2.7 Gaussiana

La principal diferencia entre la función Gaussiana y las anteriores, está en que ésta no es monótona creciente, se utiliza para extraer información relevante de los extremos. Produce salidas en el rango (0,1) y su fórmula es la siguiente:

$$F(x) = e^{-x^2} \quad (16)$$

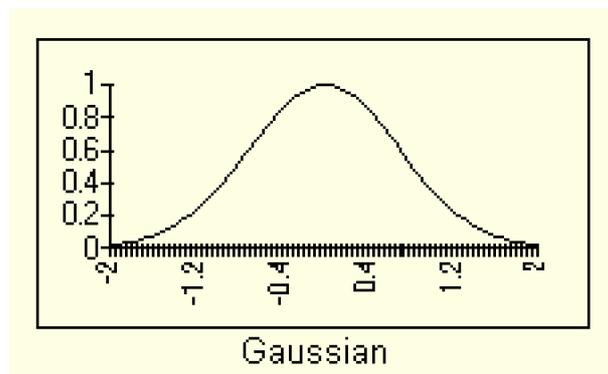


Figura 10: Función Gaussiana

## 2.8 Gaussiana Complemento

Al igual que la función Gaussiana pretende enfatizar información importante en valores extremos de los datos, con la diferencia de que para estos valores toma el valor de uno en vez de cero como en la Gaussiana (ver gráfico). Estas dos últimas funciones suelen resultar especialmente útiles al utilizarse combinadas con la función Logística o Tangente Hipérbolica en las capas ocultas de la red. Matemáticamente, esta función es:

$$F(x) = 1 - e^{-x^2} \quad (17)$$

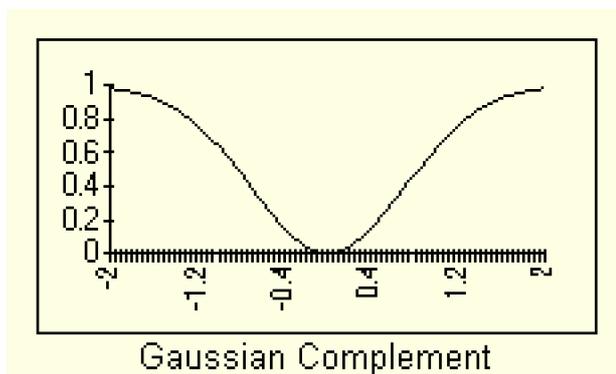


Figura 11: Función Gaussiana Complemento

### 3. Características de la Redes Neuronales Artificiales (R.N.A.)

---

Las Redes Neuronales Artificiales cuentan con varios elementos o características que las distinguen de otros sistemas. En primer lugar, su estructura se inspiró en el cerebro humano con el objeto de reproducir alguna de sus capacidades. De esta manera, están formadas por neuronas (artificiales) o elementos de procesamiento (PE), las cuales al igual que su contraparte biológica, pueden comunicarse a través de sinapsis (unión entre dos neuronas). Los elementos de procesamiento se pueden excitar o inhibir (dependiendo de la intensidad de los pesos), aprender, agruparse, etc. Cada conexión entre dos unidades tiene asignado una dirección y un peso. El conocimiento del modelo de la red queda almacenado en dichos pesos. Cada unidad de la capa oculta y de salida calcula su valor como la suma del producto de los pesos y entradas (función neta), pasando este resultado a través de una función no lineal:

$$f_i(neta_i) \quad (18)$$

A fin de resolver problemas de modo efectivo, donde se apoye la toma de decisiones en base a experiencias pasadas que se asocian y aplican en la generación de conocimiento destinado a nuevas situaciones, las Redes Neuronales se caracterizan por actuar de forma:

**a) Paralela:** Las R.N.A. cuentan con una gran cantidad de procesadores elementales, cada uno de los cuales trabaja simultáneamente con una parte de un problema mayor. Esta característica resulta de gran utilidad, por ejemplo, en el reconocimiento de imágenes visuales. Un PC que trabaja en secuencia, instrucción a instrucción, tardaría varios minutos en realizar una sencilla tarea de tratamiento sobre una imagen de 256x256 píxeles. Un sistema basado en 16 DSP operando en paralelo emplearía del orden de 20 ms en la misma tarea. En cambio, el cerebro tarda ese mismo tiempo en procesar millones de píxeles. La clave reside en que en este último caso los miles de millones de neuronas que intervienen en el proceso de visión están operando en paralelo sobre la totalidad de la imagen. <sup>16</sup>

**b) Distribuida:** Los cálculos asignados se distribuyen entre todas las neuronas, cada una de las cuales trata una fracción de un problema mayor, a diferencia de los computadores convencionales que sólo cuentan con un procesador (CPU) y memoria fija.

**c) Adaptativa:** Las R.N.A. tienen la capacidad de adaptarse al entorno modificando sus pesos y sinapsis, aprenden de la experiencia, pudiendo generalizar conceptos a partir de casos particulares y así poder encontrar una solución aceptable al problema.

## 4. Clasificación de las Redes Neuronales

Las Redes Neuronales se pueden clasificar de acuerdo a su aprendizaje, arquitectura o tipo de resultado. A continuación expondremos las distintas clasificaciones.

### 4.1 Clasificación de acuerdo a su aprendizaje

#### 4.1.1 Supervisados:

Consiste en presentarle un conjunto de patrones a la red, junto con la salida deseada u objetivo, la red es capaz de ajustar los pesos de las neuronas hasta que su salida tiende a ser la deseada, utilizando para ello información detallada del error que comete en cada paso. De esta forma, la red es capaz de estimar relaciones entrada/salida sin necesidad de proponer una cierta forma funcional de partida.

#### 4.1.2 No Supervisados:

Se le presentan a la red una multitud de patrones, sin adjuntar la respuesta que deseamos. Así, durante el proceso de aprendizaje, la red autoorganizada debe descubrir por sí misma rasgos comunes, regularidades, correlaciones o categorías en los datos de entrada, e incorporarlas a su estructura interna de conexiones (pesos). Se dice, por tanto, que las neuronas deben autoorganizarse en función de estímulos (señales o datos) procedentes del exterior. Para obtener resultados de calidad la red requiere de un cierto nivel de redundancia en las entradas procedente del espacio sensorial; en definitiva, un número de patrones de aprendizaje suficiente.

#### 4.1.3 Híbridos:

Combinación de los dos tipos básicos de aprendizaje en la red, normalmente en distintas capas de neuronas.

#### 4.1.4 Reforzados:

Se encuentran entre el aprendizaje supervisado y no supervisado. Como el supervisado, emplea información sobre el error cometido, pero solamente indicando lo bien o lo mal que se está actuando. Como en el caso no supervisado, no suministra explícitamente la salida deseada.<sup>17</sup>

<sup>16</sup> [Del Brío 97a]

## **4.2 Clasificación según su arquitectura**

### **4.2.1 Realimentadas:**

La información puede circular entre las capas en cualquier sentido, incluido el de salida-entrada. En este caso, se tiene un conjunto de neuronas que actúa como memoria de las características de los patrones procesados previamente.

### **4.2.2 Unidireccionales:**

La información fluye en un solo sentido, desde las neuronas de entrada hacia las de salida. El proceso de ajuste se lleva a cabo por épocas, existiendo un ajuste sólo en términos del error obtenido al comparar con el resultado real.

## **4.3 Clasificación según tipo de resultado**

### **4.3.1 Pronóstico o Estimación en Nivel:**

El output es un valor, una cantidad. Corresponde a una variable continua.

### **4.3.2 Clasificación:**

El output es una probabilidad, entrega categorías. Corresponde a una variable discreta.

## **5. Tipos de funcionamiento de las Redes Neuronales Artificiales**

---

### **Aprendizaje o Entrenamiento**

Cuando se construye una red neuronal, se parte de un modelo de neuronas y de una determinada arquitectura de red, estableciéndose los pesos iniciales como nulos o aleatorios. Para que la red resulte útil es necesario entrenarla, lo que se conoce como modo de aprendizaje. Aprendizaje es el proceso por el cual se produce el ajuste de los parámetros libres de la red (pesos) a partir de un proceso de estimulación por el entorno que rodea la red. En términos más simples significa la extracción de regularidades de los datos de entrenamiento que serán transferidos a nuevos ejemplos. Esto se hace a través de una cierta regla de aprendizaje (manera sistemática de cambiar los pesos), construido normalmente a partir de la optimización de una función de error o costo, que mide la eficacia actual de la operación de la red. En un proceso de aprendizaje la información contenida en los datos de entrada queda incorporada en la propia estructura de la red neuronal, estructura que almacena la representación de una cierta imagen de su entorno.

El modo de aprendizaje es de gran importancia, debido a que una característica esencial de las R.N.A. es que son sistemas entrenables, capaces de realizar un determinado tipo de procesamiento, aprendiéndolo a partir de un conjunto de ejemplos.

Hay básicamente tres aspectos relacionados con el aprendizaje: La elección del conjunto de entrenamiento y su tamaño, la elección de los parámetros del aprendizaje, y el momento en que debemos detener el aprendizaje. Para enfrentar esto sólo existen algunas reglas generales, de manera que la *experimentación* juega un papel importante. En este aspecto, la disponibilidad de ambientes de simulación rápidos y habilidades de prueba extensivos son una gran ventaja.<sup>18</sup>

## 5.2 Regla de Retropropagación

Existen diferentes algoritmos para entrenar las redes, siendo fundamental el de Retropropagación del Error, Back-propagation o Regla del Gradiente. Este algoritmo permite modificar los pesos de cada una de las neuronas de la red en forma proporcional al error de su salida (generalmente la diferencia entre lo deseado y lo obtenido), el cual por medio de la regla se va propagando desde la salida final por las capas intermedias hasta la entrada. Los pasos a seguir para este entrenamiento son los siguientes:

- a) Aplicar un vector de entrada como estímulo para la capa de entrada.
- b) Propagar esta señal por todas las capas de la red hasta obtener una salida.
- c) Comparar la salida de la red con el output deseado.
- d) Calcular una señal de error para cada unidad de salida.
- e) Transmitir la señal de error hacia atrás hacia todos los nodos que estén conectados directamente.
- f) Los nodos de capas anteriores reciben una fracción del error proporcional al peso de su conexión.
- g) Actualizar iterativamente los pesos de todas las conexiones de tal manera de disminuir el error.
- h) En el caso de que no se tenga convergencia, repetir todo lo anterior. Cuando el error resulta aceptablemente pequeño, el aprendizaje se da por concluido.<sup>19</sup>

## 5.3 Elección del conjunto de entrenamiento

El tamaño del conjunto de entrenamiento es de vital importancia para la utilidad práctica de la red. Si los patrones de entrenamiento no transfieren todas las características del problema, el mapeo descubierto durante el entrenamiento sólo se aplica al conjunto de entrenamiento. Así, el rendimiento en el conjunto de prueba será mucho peor que el rendimiento del conjunto de entrenamiento. La única regla general es usar muchos datos y representativos, es decir, datos que cubran todos los escenarios posibles con los que se pueda encontrar la red en el futuro.

La relación entre el tamaño y el conjunto de entrenamiento y el número de pesos de la red neuronal juega un papel importante. Si el número de ejemplos de entrenamiento es

<sup>18</sup> [Reyes 02e]

<sup>19</sup> [Virreira 96d]

menor que el número de pesos se podría producir una memorización de la solución, es decir, que la red almacene cada ejemplo del conjunto de entrenamiento, lo que se traducirá en una pobre generalización (capacidad de proporcionar una respuesta correcta ante datos no utilizados en su entrenamiento). Es recomendable que el número de muestras de entrenamiento sea de al menos el doble que el número de pesos de la red.

Cuando hay una gran discrepancia entre el rendimiento del conjunto de entrenamiento y el de prueba, se debe detener el aprendizaje. En estos casos se debe aumentar el tamaño del conjunto de entrenamiento y/o producir una mezcla diferente de ejemplos de entrenamiento y prueba. Siempre se puede esperar una caída en el rendimiento (más o menos entre 10% a 15%) desde el conjunto de entrenamiento al de prueba.<sup>20</sup>

#### **5.4 Tamaño de la red**

La forma más eficiente para determinar el tamaño óptimo de la red es a través de la *experimentación*. Un parámetro importante es el número de neuronas en la capa oculta, el cual está relacionado con la capacidad de mapeo de la red. Mientras más grande es su número, mayor es su capacidad para memorizar el conjunto de entrenamiento. Sin embargo, si se continúa aumentando el tamaño de la red, hay un punto donde la generalización empeorará, debido al hecho que se puede sobreajustar el conjunto de entrenamiento, de modo que cuando la red trabaja con patrones que nunca ha visto anteriormente la respuesta es impredecible. El problema es encontrar cuál es la combinación correcta de neuronas y número de capas para resolver el problema en tiempos de entrenamiento más bajo. Cada capa aumenta el poder discriminante de la red. Es recomendable comenzar con una red pequeña y observar el comportamiento de la curva de aprendizaje. Si el error final es pequeño, el número de procesadores elementales probablemente es el apropiado. Si el error final es grande, el aprendizaje puede haber caído en un mínimo local (con lo cual el número de neuronas debe ser aumentado). En el caso que el rendimiento de la red se deteriore considerablemente desde el conjunto de entrenamiento al conjunto de prueba, una de dos cosas ha ocurrido: el conjunto de entrenamiento no es representativo del dominio del problema, o se ha configurado la red con muchos pesos produciendo una pobre generalización (sobreajuste).

Un método alternativo para el tamaño de la red, es comenzar con una red grande e ir cambiando alguno de los pesos. Los pesos pequeños que tiendan a cero y decaen de iteración a iteración, serán eliminados.<sup>21</sup>

#### **5.5 Parámetros de aprendizaje**

No existe una recomendación o receta en lo que se refiere a parámetros de aprendizaje. El objetivo es entrenar tan rápido como sea posible y alcanzar el mejor rendimiento.

---

<sup>20</sup> [Reyes 02f]

<sup>21</sup> [Reyes 02g]

Aumentando el parámetro de la tasa de aprendizaje, se disminuirá el tiempo de entrenamiento, pero también aumentarán las posibilidades de divergencia, es decir, de alejarse del valor óptimo. Debido a que la corrección del peso depende de las características de la superficie de error y de la tasa de aprendizaje, para obtener un aprendizaje constante es necesario un parámetro adaptativo. Se recomienda una estrategia en donde la tasa de aprendizaje sea grande al comienzo del proceso de aprendizaje y que progresivamente decaiga hacia el término de la adaptación.

Las actualizaciones pueden ejecutarse al término de la presentación de todos los elementos del conjunto de entrenamiento o en cada iteración (tiempo real). La primera modalidad suaviza el gradiente y puede dar un aprendizaje más rápido para datos ruidosos, sin embargo, también puede promediar los gradientes en cero y evitar el aprendizaje. La modificación de los pesos en cada iteración con una tasa de aprendizaje pequeña puede ser preferible la mayor parte del tiempo.

La curva del ECM (error cuadrático medio) como función del número de iteraciones es llamada *curva de aprendizaje*. Cuando la curva de aprendizaje presenta una forma horizontal, el tamaño del paso (parámetro que determina la magnitud en la cual se van a actualizar los pesos) debería ser incrementado para acelerar el aprendizaje y mejorar su rendimiento. Por otro lado, cuando la curva de aprendizaje presenta una forma oscilatoria el tamaño del paso debiera ser disminuido. En un caso extremo, el error aumenta uniformemente, mostrando que el aprendizaje es inestable. En este punto la red debería ser reformulada. Cuando la curva de aprendizaje se estabiliza después de muchas iteraciones en un nivel que no es aceptable, es tiempo de cambiar la topología de la red (más neuronas en la capa oculta o más capas ocultas, o una topología diferente) o el procedimiento de entrenamiento.<sup>22</sup>

### 5.6 Criterio de detención

Otro problema es cuándo detener el entrenamiento. Todos los criterios están basados en el error cuadrático medio. Los criterios más utilizados son fijar el número de iteraciones o el prefijar un error final. Lo que se usa en la práctica es el aprendizaje mínimo incremental (cuando entre dos iteraciones consecutivas el error no disminuye en al menos una cantidad dada o umbral, el entrenamiento debiera terminarse). Esto nos da un criterio para comparar topologías muy diferentes. Otra alternativa es usar dos criterios para detener el entrenamiento: fijar un ECM para el conjunto de prueba o fijar un número máximo de eventos (es la presentación de un solo patrón de entrenamiento a la red) desde que se alcanzó el último mínimo. El aprendizaje se detiene cualquiera de las dos alternativas anteriores es alcanzada.<sup>23</sup>

### 5.7 Memorización v/s Generalización

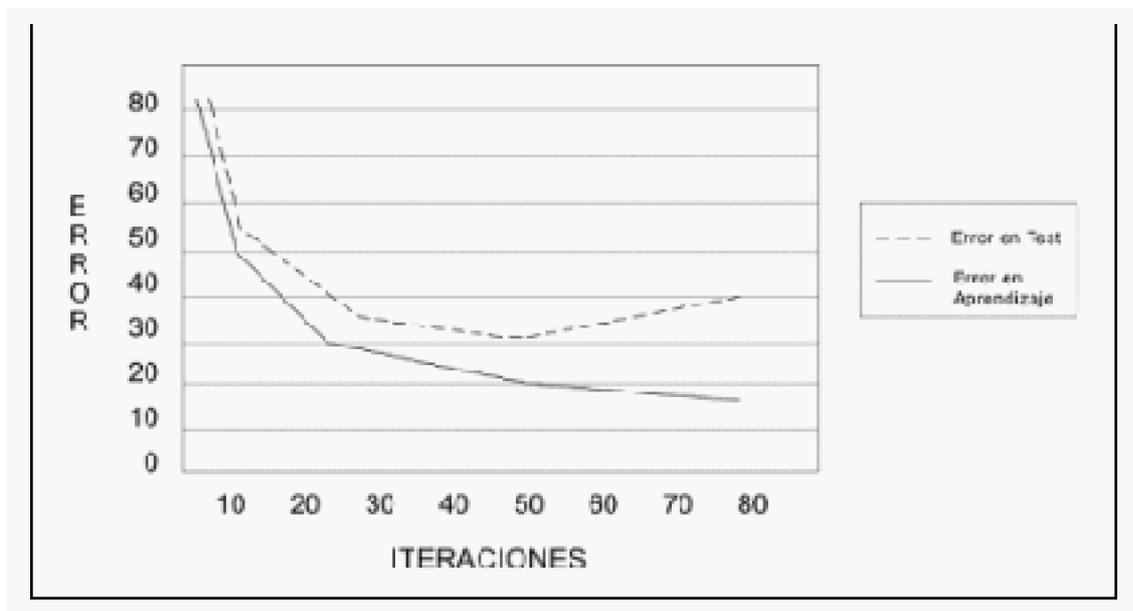
Uno de los aspectos fundamentales se relaciona con la capacidad de las redes

<sup>22</sup> [Reyes 02h]

<sup>23</sup> [Reyes 02i]

neuronales de generalizar a partir de ejemplos, lo que se denomina en ocasiones el problema de la *memorización* frente a la *generalización*. La Generalización es la capacidad de proporcionar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento. En su proceso de entrenamiento se puede considerar, por una parte, un *error en aprendizaje* y por otra un *error en generalización* (error en test). Una red se debe entrenar hasta que su *error de generalización* sea mínimo.

Para medir de una forma objetiva la eficacia final del sistema construido, el conjunto de patrones de partida debería ser dividido en tres grupos, un conjunto de *aprendizaje* y dos de test, uno para la realización de *validación cruzada* (que evite el sobreajuste) y otro que se mantenga al margen del proceso de aprendizaje, permitiendo de este modo una validación totalmente objetiva de la red neuronal desarrollada.



*Error en Test v/s error en Aprendizaje*

### **5.8 Recuerdo o Ejecución**

Una vez terminado el aprendizaje, los pesos y la estructura quedan fijos, la red neuronal está en condiciones de procesar datos. Esto es lo que se conoce como *modo recuerdo o ejecución*.

## **6. Tipos de Redes Neuronales Artificiales**

---

### **6.1 Perceptrón Multicapa (MLP)**

Son redes unidireccionales entrenadas con el algoritmo de propagación hacia atrás (BP). Son redes con aprendizaje supervisado, de manera que requieren una respuesta deseada para ser entrenadas. Son ampliamente usados para clasificar patrones. Usan eficientemente la información contenida en los datos de entrada. Estos modelos actúan

como funciones mapeadoras universales, desempeñándose muy bien con las series de tiempo.<sup>24</sup>

La red aprende un conjunto predefinido de pares de entrada y salida dados como ejemplos, empleando un ciclo de propagación-adaptación de dos fases: una vez que se han aplicado los datos de entrada a la primera capa de unidades de la red, ésta se va propagando a través de todas las capas superiores hasta generar una salida, la cual se compara con la salida deseada para estimar el error asociado a la predicción. El error se transmite hacia atrás, desde la capa de salida hacia todas las capas intermedias que contribuyan directamente hacia ella. Sin embargo, las unidades de la capa intermedia sólo reciben una fracción del error total, basándose en la contribución relativa de la unidad a la salida original. Este proceso se repite, capa por capa. Luego, basándose en el error, se actualizan los pesos de conexión de cada unidad con el objeto de ajustar el modelo y disminuir el error final. Así, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que ellas aprenden a reconocer distintas características del conjunto de entrenamiento, aproximándose de esta manera al pensamiento y al comportamiento de los agentes, en nuestro caso, del mercado cambiario.

Según Martín del Brío y Sanz (1997), el aprendizaje se puede dar por concluido cuando el error resulta aceptablemente pequeño para todos los pares de vectores de entrenamiento, es decir, cuando se logra minimizar el error de la estimación. Como señalamos anteriormente, el término de error que emplean los *softwares* de redes neuronales es el error cuadrático medio. No obstante, la medida de error empleada en este estudio estará dada por la prueba de signo ya que, de acuerdo a Leung, Daouk y Chen (2000), los modelos propuestos para predecir el signo se desempeñan mejor que los modelos de proyección de nivel en términos de su tasa de acierto (medida por el número de veces en que la dirección pronosticada es correcta) y, además, son capaces de generar beneficios más altos.<sup>25</sup>

### **6.1.1 Ventajas del MLP**

a) Se puede usar en una multitud de problemas diferentes, proporcionando con frecuencia buenas soluciones con no demasiado tiempo de desarrollo.

b) Con una o dos capas ocultas pueden aproximar virtualmente cualquier mapa de entrada-salida.

### **6.1.2 Desventajas del MLP**

a) Su lentitud de convergencia. Para este inconveniente existe una variante a modo de corrección del problema. Se trata de incluir un término adicional al algoritmo de aprendizaje (BP), el denominado término de *momento*, que consiste en añadir al cálculo de la variación de los pesos, un término adicional proporcional al incremento de la

<sup>24</sup> [Parisi 02a]

<sup>25</sup> [Parisi 02b]

iteración anterior. Si el incremento en un determinado peso es alto, en la siguiente iteración lo será todavía más. Por otra parte, si los incrementos en un determinado peso oscilan (a veces son positivos, otras negativos), el incremento efectivo acumulado se reduce al cancelarse. Esta es una manera de aumentar el ritmo de aprendizaje efectivo en determinadas direcciones.

Puede incurrir en sobreajuste.

b) No garantiza alcanzar el mínimo global de la función de error.

c) Son clasificadores estáticos, es decir, el mapa de entrada-salida depende sólo de la entrada actual.

d) Necesita un gran conjunto de datos de entrada.

## **6.2 Variantes del MLP**

### **6.2.1 Conexión Jump**

En este tipo de red de retropropagación, cada capa está conectada a la o las capas anteriores, es decir, por ejemplo en una red de tres capas, la capa oculta está conectada a la de entrada y la capa de salida lo está con la de entrada y la oculta. Esta arquitectura puede ser útil cuando se está trabajando con problemas muy complejos, es decir, cuando puede ser muy difícil para una persona definir los diferentes patrones que son inherentes a los datos.

### **6.2.2 MLP con Conexiones Recurrentes**

El MLP con conexiones recurrentes es conocido por su habilidad para aprender secuencias y, por lo tanto, es recomendable para series de tiempo. Las redes con algoritmo de propagación hacia atrás y conexiones recurrentes son entrenadas de la misma forma que la red BP estándar, excepto que los patrones deben ser presentados siempre en el mismo orden y la selección aleatoria no está permitida, por el hecho de operar con series de tiempo. La única diferencia en estructura es que existe un *slab* (grupo de neuronas del mismo tipo) extra en la capa de entrada que está conectado a la capa oculta, al igual que el otro *slab* de entrada, lo que le permite a la red ver el conocimiento anterior que tenía sobre las entradas previas y, de esta manera, operar con rezagos de las variables explicativas.<sup>26</sup>

Existen 3 tipos de redes MLP con conexiones recurrentes:

a) La capa de entrada realimentada en la capa de entrada: el término largo de memoria (“slab extra”), recuerda los nuevos datos de entrada y los usa cuando el próximo patrón es procesado.

---

<sup>26</sup> [Parisi 02c]

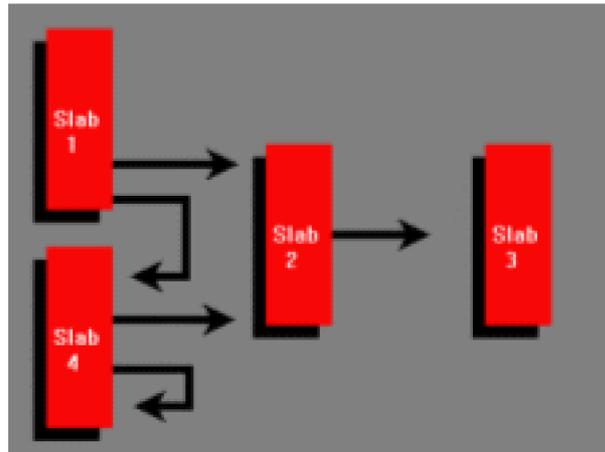


Figura 13: Capa de entrada realimentada en la capa de entrada

b) La capa oculta realimentada en la capa de entrada: el término largo de memoria recuerda la capa oculta, la cual contiene las características detectadas en los datos de patrones anteriores. Esta es la red recurrente más poderosa.

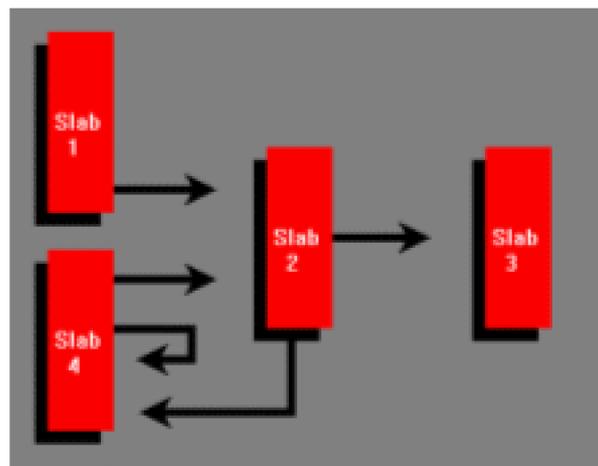


Figura 14: La capa oculta realimentada en la capa de entrada

c) La capa de salida realimentada a la capa de entrada: el término largo de memoria recuerda la salida previamente predicha.

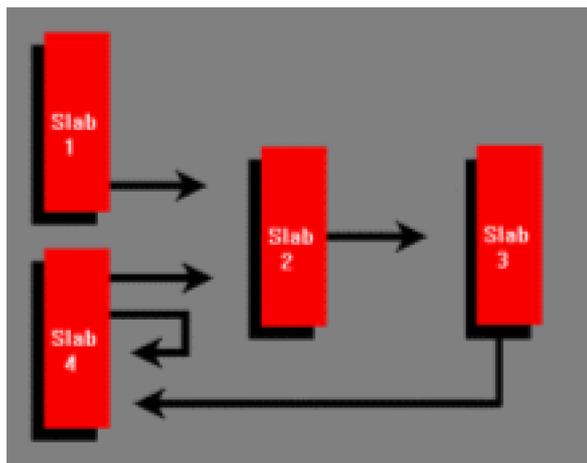


Figura 15: Capa de salida realimentada a la capa de entrada

### 6.2.3 Red Ward

La red Ward se caracteriza porque pueden aplicarse diferentes funciones de activación a los *slabs* de la capa oculta, para detectar diferentes características en los patrones procesados a través de la red. Por ejemplo, un diseño de una red puede usar una función de activación *Gaussiana* en un *slab* oculto para detectar diferentes características en el rango medio de los datos, y usar un *Complemento Gaussiano* en otro *slab* oculto para detectar características de los extremos superiores e inferiores de los datos. De esta manera, la capa de salida tendrá diferentes puntos de vista de los datos, lo cual puede conducir a una mejor predicción.

Existen 3 tipos de red Ward:

a) Dos *slabs* ocultos con diferentes funciones de activación: es una red BP regular de tres capas con dos *slabs* en la capa oculta. Usa una función de activación diferente para cada *slab* en la capa oculta para detectar diferentes características en los datos.

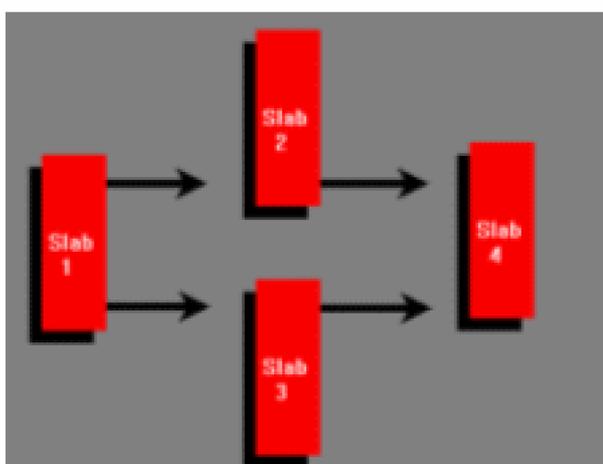


Figura 16: Dos *slabs* ocultos con diferentes funciones de activación

b) Tres *slabs* ocultos con diferentes funciones de activación: es una red BP que

agrega un tercer *slab* a la capa oculta. Cuando cada *slab* en la capa oculta tiene una función de activación diferente, ofrece tres formas diferentes de ver los datos.

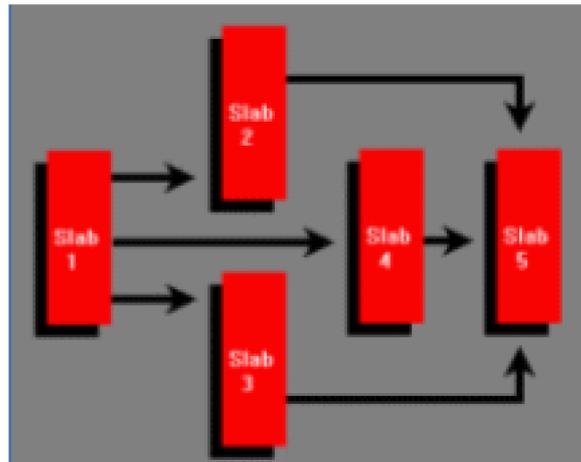


Figura 17: Tres slabs ocultos con diferentes funciones de activación

c) Dos *slabs* ocultos, diferentes funciones de activación más una conexión Jump: esta red Ward es un BP regular con dos *slabs* en la capa oculta y una conexión Jump entre la capa de entrada y la capa de salida. La capa de salida recibe dos vistas diferentes de los rasgos de los datos como descubrió en los *slabs* ocultos más las entradas originales.

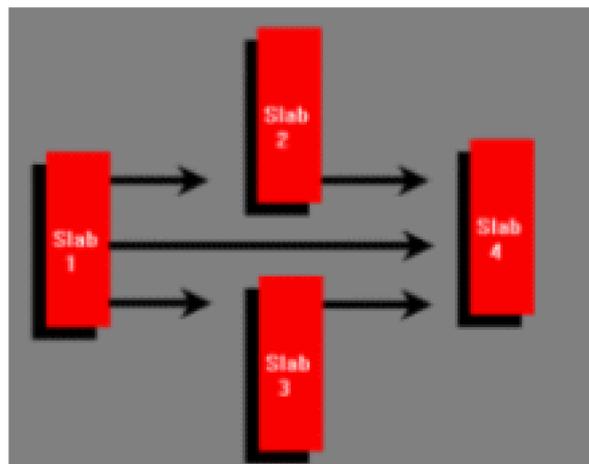


Figura 18: Dos slabs ocultos, diferentes funciones de activación más una conexión Jump

### 6.3 Función de Base Radial

Es un modelo híbrido que incorpora tanto aprendizaje supervisado como no supervisado. En su estructura cuenta con 3 capas de neuronas:

a) Capa de entrada: Donde las neuronas simplemente envían la información del exterior hacia las neuronas de la capa oculta.

b) Capa oculta: Las neuronas de la capa oculta operan en base a la distancia que

separa el vector de entrada respecto del vector sináptico que cada una almacena (denominado Centroides) y su respuesta es localizada, pues sólo responden con una intensidad apreciable cuando el vector de entradas presentado y el Centroides de la neurona pertenecen a una zona próxima en el espacio de las entradas.

c) Capa de salida: En esta capa, las neuronas son lineales, esencialmente calculan la suma ponderada de las salidas que proporciona la capa oculta.

### **6.3.1 Ventaja**

Encuentra la entrada al mapa de salida usando aproximadores locales. Se entrena rápidamente y requiere pocas muestras de entrenamiento.

### **6.3.2 Desventaja**

Un problema es que puede requerir un espacio dimensional muy grande para solucionar un problema.

## **6.4 Análisis de Componentes Principales (PCA)**

Es una mezcla de aprendizaje supervisado y no supervisado. El PCA es un procedimiento lineal para encontrar la dirección en el espacio de entrada en donde se produce la mayor parte de la energía de la entrada. En otras palabras, el PCA ejecuta una extracción de características. Las proyecciones de esos componentes corresponden a los valores propios de la matriz Covarianza de Entrada. El segmento no supervisado de la red ejecuta la extracción de características y el segmento supervisado de la red ejecuta la clasificación (lineal o no) de esas características usando un MLP.

Esta técnica transforma las muestras de entrada en un nuevo espacio de modo que la información acerca de las muestras es mantenida, pero la dimensionalidad reducida. Esto hace que la clasificación trabaje en forma más fácil.

La importancia del análisis PCA es que el número de entradas para el MLP puede ser significativamente reducido. Esto conlleva a una reducción del número de patrones de entrenamiento requerido y una reducción en el tiempo de entrenamiento.<sup>27</sup>

### **6.4.1 Ventajas**

a) Converge rápidamente.

b) El segmento no supervisado provee un medio de reducción de datos.

### **6.4.2 Desventaja**

Posibilidad de perder importante información de entrada.

## **6.5 Mapas Autoorganizados de Características (SOFM)**

<sup>27</sup> [Reyes 02]

En este modelo, las neuronas se organizan en una arquitectura unidireccional de dos capas. La primera es la capa de Entrada o Sensorial, la cual distribuye la información procedente del espacio de entrada a las neuronas de la segunda capa, en donde se lleva a cabo el procesamiento, formando así el mapa de rasgos. Es una red no supervisada con comportamiento “competitivo”.

En el modo de operación normal de la red, los pesos permanecen fijos. En primer lugar, cada neurona  $(i,j)$  calcula la similitud entre el vector de entradas  $\mathbf{X}$ , y su propio vector de pesos sinápticos  $W_{ij}$  según una cierta medida de distancia o criterio de similitud establecido. A continuación, se declara vencedora la neurona  $g$  cuyo vector de pesos  $W_g$  es más similar al de entradas.

En la fase de *aprendizaje* cada neurona del mapa sintoniza con diferentes rasgos del espacio de entrada. El proceso es el siguiente. Después de la presentación y procesamiento de un vector de entrada  $\mathbf{X}(t)$ , la neurona vencedora modifica sus pesos de manera que se parezcan un poco más a  $\mathbf{X}(t)$ . De este modo, ante el mismo patrón de entrada, dicha neurona responderá en el futuro todavía con más intensidad.

Lo descrito hasta el momento responde a un esquema competitivo clásico, en el que cada neurona actúa en solitario. Sin embargo, el modelo de mapa de Kohonen aporta una importante novedad, ya que incorpora a este esquema relaciones entre las neuronas próximas al mapa. Para ello, introduce una denominada *función de vecindad*, que define un entorno alrededor de la neurona ganadora actual (vecindad). Su efecto es que durante el aprendizaje se actualizan tanto los pesos de la vencedora como los de las neuronas pertenecientes a su vecindad. De esta manera, en el modelo de SOFM se logra que neuronas próximas sintonicen con patrones similares.<sup>28</sup>

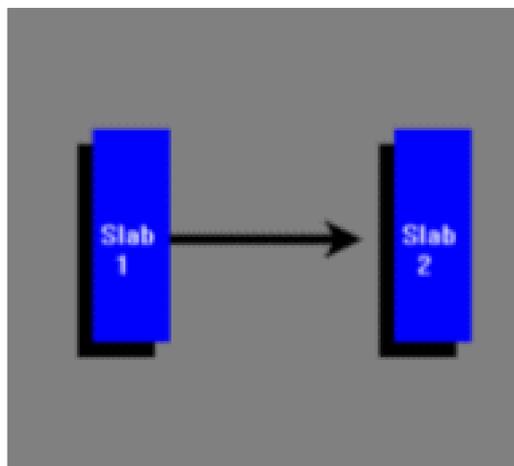


Figura 19: Red SOFM

La utilización de la función vecindad en el modelo de mapas auto-organizados aporta en comparación al modelo competitivo sencillo dos ventajas adicionales: el ritmo efectivo de convergencia se mejora y el sistema es más fuerte frente a variaciones en los valores iniciales de los pesos.

<sup>28</sup> [Larraín 90a]

La principal novedad de los SOFM consiste en que la modificación de los pesos no sólo se aplica a una neurona específica (la ganadora), sino también a su vecindad. Así, el proceso de aprendizaje comprende dos fases fundamentales: una ordenación global, en la que se produce el despliegue del mapa, y un ajuste fino en el que las neuronas se especializan.

### **6.5.1 Ventajas**

a) Reduce el espacio de entrada en características representativas usando un proceso de autoorganización, de manera que el espacio de entrada es mantenido, mientras que la dimensionalidad del espacio es reducido.

b) Tiene la capacidad de agrupar y clasificar la información.

### **6.5.2 Desventaja**

Es difícil medir su rendimiento, lo que puede afectar la estabilidad del mapa.

## **6.6 Time Lagged Recurrent Network (Redes Recurrentes con Retardo Temporal, TLRN)**

Son MLP extendidos con conexiones recurrentes locales. Es un modelo muy apropiado para el procesamiento de información temporal, la predicción de series de tiempo, sistemas de identificación y reconocimiento de patrones temporales. El algoritmo de entrenamiento usado es el *backpropagation through time* (BPTT), mecanismo de aprendizaje dinámico que permite especificar ejemplares compuestos por una cantidad de muestras equivalentes a una secuencia de tiempo. Estas muestras son usadas para retropropagar el error a través del tiempo.

Los tipos de memoria disponibles para el TLRN son:

**a) Tap Delay Neuronal Networks (TDNN):** Consta de una gran capacidad para imitar el comportamiento de la serie. La memorización alcanza un gran grado de detalle. La estructura de memoria TDNN es simplemente una cascada de retardos ideales (retardos de una muestra).

**b) Gamma:** Provee una memoria recursiva de las señales de entradas pasadas.

**c) Laguerre:** Es ligeramente más sofisticada que la gamma, ya que ortogonaliza el espacio de memoria, lo cual es útil cuando la red trabaja con núcleos grandes de memoria.<sup>29</sup>

### **6.6.1 Ventajas**

a) Requiere una red pequeña para aprender problemas temporales.

b) Posee una menor sensibilidad al ruido.

---

<sup>29</sup> [Jiménez 96b]

### 6.6.2 Desventajas

- a) El entrenamiento puede caer en un mínimo local.
- b) El BPTT es muy complejo y requiere mucha memoria.

### 6.7 Red Neuronal Probabilística (PNN)

Estas redes son conocidas por su habilidad para entrenar rápidamente un conjunto amplio de datos. La PNN separa datos en un número especificado de categorías de salidas.

La red PNN trabaja comparando el nuevo patrón de entrada y cada uno de los patrones de entrenamiento de acuerdo a su distancia.

La PNN es una red de tres capas en donde los patrones de entrenamiento son presentados a la capa de entrada y la capa de salida tiene una neurona para cada posible categoría. Deben existir tantas neuronas en la capa oculta como patrones de entrenamiento.

La red produce activaciones en la capa de salida correspondiente a la estimación de la función de densidad de probabilidad para esa categoría. La salida más alta representa la categoría más probable.

El número de neuronas en la capa de entrada es el número de entradas en el problema, y el número de neuronas en la capa de salida corresponde al número de categorías. Debido a que el propósito de una red PNN es separar salidas en categorías diferentes, dos o más salidas son requeridas. Los valores de salida son cero o uno.

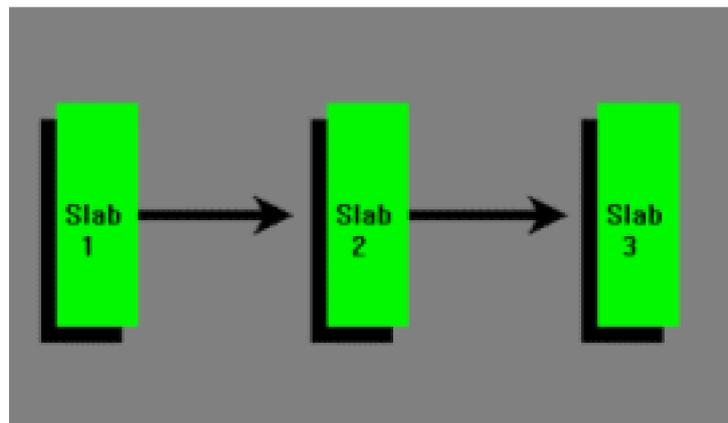


Figura 20: Red Neuronal Probabilística

### 6.8 Red Neuronal de Regresión General (GRNN)

Estas redes son conocidas por su habilidad para entrenar rápidamente un conjunto amplio de datos. Es un tipo de red supervisada. La GRNN trabaja midiendo cuán lejos un patrón de muestra dado está de los patrones en el conjunto de entrenamiento en un espacio dimensional  $N$ , donde  $N$  es el número de entradas en el problema. Cuando un

nuevo patrón es presentado a la red, ese patrón de entrada es comparado en un espacio dimensional N a todos los patrones en el conjunto de entrenamiento para determinar cuan lejos en distancia está de aquellos patrones. La salida que es predecida por la red es una cantidad proporcional de todas las salidas en el conjunto de entrenamiento. La proporción está basada en cuan lejos el nuevo patrón está de los patrones dados en el conjunto de entrenamiento.

Al igual que la red PNN categoriza datos, pero tiene la capacidad para producir valores de salida continua. Es especialmente útil para funciones de aproximación continuas. La GRNN puede tener entradas multidimensionales, y ajustará un espacio multidimensional a través de los datos.

La GRNN es una red de tres capas donde debe existir una neurona oculta para cada patrón de entrenamiento. No hay parámetros de entrenamiento tales como tasa de aprendizaje y momentos como en el BP, pero hay un factor *smoothing* que es aplicado después que la red es entrenada.<sup>30</sup>

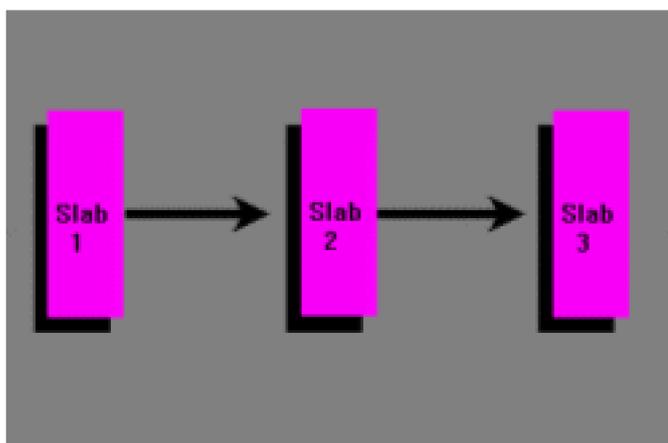


Figura 21: Red Neuronal de Regresión General (GRNN)

### 6.9 Red GMDH (Group Method of Data Handling)

La red GMDH trabaja construyendo capas sucesivas con uniones (o conexiones) complejas que son los términos individuales de un polinomio. Estos términos polinomiales son creados usando regresión lineal y no lineal. La capa inicial simplemente es la capa de entrada. La primera capa creada es hecha calculando regresiones de los valores en la primera capa junto con las variables de la entrada. De nuevo, sólo los mejores son escogidos por el algoritmo. Éstos se llaman *sobrevivientes*. Este proceso continúa hasta que la red se detiene obteniendo los mejores (según un criterio de selección especificado).

La red resultante puede representarse como un polinomio complejo (es decir, una fórmula familiar). Se puede ver la fórmula que contiene las variables de la entrada más significativas. En algunos aspectos, es como usar análisis de la regresión, pero es más

<sup>30</sup> [Reyes 02k]

poderoso que el análisis de la regresión. La red GMDH puede construir modelos muy complejos evitando problemas de sobreajuste.

La red GMDH contiene varios métodos de evaluación, llamados criterios de selección, para determinar cuándo debiera terminar el entrenamiento.

Una ventaja de la red GMDH es que reconoce las variables más significativas a medida que entrena, y desplegará una lista de ellas.

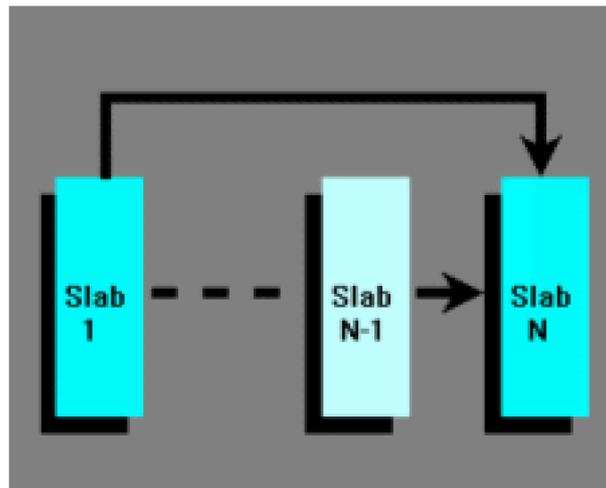


Figura 22: Red GMDH (Group Method of Data Handling)

## 7. Ventajas de las Redes Neuronales Artificiales

Las mayores ventajas de procesamiento de los sistemas de Redes Neuronales Artificiales son:

**a) Adaptabilidad:** Esta es la capacidad de las redes neuronales de aprender cómo realizar ciertas tareas al presentarle ejemplos ilustrativos. Si es necesario, una red neuronal ya entrenada y operando puede realizar un aprendizaje continuo presentándole constantemente nuevos ejemplos. En este sentido, las redes neuronales pueden mejorar su capacidad de discriminar entre entradas similares o adaptarse a cambios del ambiente.

**b) Tolerancia a fallas:** Si una neurona individual o una porción completa de la red neuronal es destruida, continúa funcionando adecuadamente. Cuando el daño llega a ser tan extensivo que el comportamiento de la red comienza a ser afectado, el efecto es una degradación gradual de rendimiento, en forma opuesta a una falla completa.

La razón por la cual las redes neuronales son capaces de tener este tipo de tolerancia a fallas, es que ellas almacenan la información de una manera distribuida. Una pieza de conocimiento o información no es codificada en una sola neurona sino que sobre varias. Esto contrasta con las tecnologías tradicionales de procesamiento de la información, en las cuales cada elemento de información es almacenado en una localización específica de memoria. En estos casos, un pequeño mal funcionamiento de *hardware* puede causar pérdidas catastróficas de información y fallas de programas.

**c) Autoorganización:** Durante el entrenamiento, una red neuronal organiza y crea su propia representación de la información que recibe. Esto es algo diferente a las metodologías tradicionales de programación, las cuales requieren que toda la información sea especificada explícitamente por el programador. Consecuentemente, las redes neuronales son muy adecuadas para problemas en los cuales es difícil o imposible definir un modelo explícito, programas o reglas para obtener la solución. Esta es una ventaja clara, ya que existe un gran número de problemas de este tipo.

Un resultado útil de esta capacidad de autoorganización es que la red neuronal logra la habilidad de generalización. Una red neuronal entrenada para reconocer ciertos patrones creará su propia categoría de patrones levemente diferente o poco familiar intentará asociarlo con alguna de las categorías de patrones que ya ha aprendido.

**d) Procesamiento rápido:** Uno de los intereses más críticos en las tecnologías de procesamientos de datos ha sido siempre la velocidad de procesamiento. A partir de intentos de automatizar procesos que envuelven grandes volúmenes de información y de gran complejidad, ha surgido la necesidad cada vez más urgente, de desarrollar programas y equipos de alta eficiencia en el procesamiento de datos.

Tradicionalmente, las investigaciones en cuanto a eficiencia computacional se han dirigido primariamente al desarrollo de programas seriales y equipos que los hagan funcionar en forma rápida, por un lado mediante el aumento de eficiencia de los programas, a través del diseño de procesadores con arquitecturas mejoradas; y por otro, mejorando la capacidad y manejo de la memoria computacional.

Las redes neuronales han apuntado a paralelizar el procesamiento de datos mediante la división de tareas computacionales en pequeñas subtareas, las cuales son repartidas sobre varios procesadores que operan simultáneamente.

**e) Compatibilidad con las tecnologías existentes:** Al igual como cualquier nueva idea o tecnología, un nuevo método de procesamiento como las redes neuronales tendría muy poca oportunidad de éxito y aceptación si fuera incompatible y difícil de integrar con otras tecnologías establecidas. Las redes neuronales son prácticas en este aspecto ya que pueden ser fácilmente implementadas utilizando técnicas existentes, y fácilmente integradas a casi cualquier tipo de sistema.<sup>31</sup>

## **8. Desventajas de las Redes Neuronales**

---

Las Redes Neuronales Artificiales son modelos que ofrecen numerosas ventajas, pero también presentan algunos inconvenientes:

a) Las R.N.A. constituyen un método de resolución de problemas demasiado creativo, es decir, que dada las especificaciones de un problema, se desconoce la topología con la que se va a solucionar del modo más eficiente. Para ello hay que utilizar el ya conocido método de prueba y error.

b) Una vez entrenada una red neuronal, se hace difícil interpretar su funcionamiento.

---

<sup>31</sup> [Jiménez 96c]

Aún más no es fácil asegurar con que grado de acierto responderá ante casos nunca vistos (generalización).

c) Es difícil averiguar por qué una red neuronal no es capaz de ajustar los datos que le son proporcionados. Esto implica que dado un problema, puede resultar dificultoso saber por qué no funciona correctamente, ya que existen una multitud de factores que pueden influir en su rendimiento.

d) Los modelos neuronales necesitan una herramienta de procesamiento poderosa. Esto se manifiesta principalmente en el proceso de aprendizaje, pero este inconveniente se contrarresta con la facilidad de implementación en dispositivos de *hardware* específicos.



## IV. MODELOS DE PREDICCIÓN DE REDES NEURONALES

La necesidad de contar con técnicas que permitan predecir adecuadamente la evolución futura del valor del dólar y que ayuden a las personas o inversionistas a tomar una mejor decisión respecto a la compra y venta de dólares ha dado origen a diferentes técnicas de análisis. En este estudio podemos comprobar como una de las técnicas más importantes a nivel de predicción son las Redes Neuronales.

Cabe señalar que el principal objetivo de todos estos análisis es el de ser instrumentos de apoyo para el inversionista, ya que le entregan una mejor orientación, ayudándole a una buena toma de decisiones, determinando el momento óptimo para comprar o vender divisas.

Hemos decidido predecir la dirección de la fluctuación del tipo de cambio, antes que el nivel de precios ya que como han demostrado Leung, Daouk & Chen y que posteriormente fueron apoyados por O'Connor, Remous & Griggs (1997), la capacidad predictiva de los modelos de clasificación es mayor a la de los modelos con estimación de nivel, en términos de su tasa de acierto y son capaces de generar beneficios más altos. Cabe destacar que Hodgson & Nicholls (1991) sugieren evaluar la significancia económica de predecir la dirección de los cambios en los precios de los activos y no su nivel.<sup>32</sup>

---

<sup>32</sup> [Parisi 02d]

Las redes neuronales pueden entenderse como modelos multiecuacionales (o multi-etapas), simultáneas y/o secuenciales, en los que el output de unas constituye el input de otras. Pueden haber múltiples inputs (variables explicativas) y múltiples outputs (proyecciones de diferentes variables). Entre los inputs y los outputs hay una capa (o múltiples capas) de procesamiento que imita el trabajo del cerebro humano.<sup>33</sup>

Las redes neuronales, nos ayudan a predecir el comportamiento de las personas a partir de datos que reflejan cómo se han comportado anteriormente. Las personas somos capaces de incorporar información en nuestro cerebro, y muchas veces lo hacemos de diversas fuentes y poseemos la capacidad de integrar toda esta información para luego aprender de ésta, encontrar patrones y extraer conclusiones que nos van a llevar a tomar decisiones en nuestra vida. Es así como las R.N.A. tratan de imitar todos estos “pasos” en el procesamiento de información que realizamos al tomar decisiones. Primero, se le entregan inputs al programa (Neuroshell), que va a ser la información con la cual se va trabajar. Ésta es procesada con el propósito que el programa pueda obtener algún grado de aprendizaje y después lo aplique en un conjunto de datos que no ha estudiado, para así poder obtener predicciones.

Para lograr obtener estos resultados vamos a utilizar dos técnicas de funcionamiento: recursivo y rolling aplicado en las redes neuronales. El funcionamiento rolling le otorga una mayor importancia a los datos que se tienen más recientemente y no toma en cuenta los datos más antiguos. El funcionamiento recursivo agrega nueva información, pero a la vez mantiene la que se tiene anteriormente.

Lo que buscamos en este trabajo es encontrar modelos de redes neuronales para la predicción de la variación del precio del dólar y comparar el funcionamiento del sistema recursivo versus el sistema rolling. Lo que queremos predecir va a ser el movimiento que se observa en el tipo de cambio.

## **A. METODOLOGÍA**

### **1. Variables de entrada**

---

Con el fin de lograr mejores y más certeros resultados en nuestra predicción, ocupamos el software Neuroshell, el cual necesita por lo menos que todos los datos se dividan en dos grandes grupos. Esto se hace para poder ordenar el funcionamiento del programa a través de un aprendizaje con el primer grupo de datos (conjunto de entrenamiento) y realice una aplicación de lo aprendido en el segundo grupo (conjunto extramuestral).

Las variables que escogimos son las siguientes:

#### **1.1 Dow Jones Industry**

---

<sup>33</sup> [Parisi 02e]

Los promedios del Dow Jones tienen su origen en la iniciativa de varios editores financieros: Charles Henry Dow (1851-1902) y de Edward D. Jones (1856-1920). Ambos periodistas, quienes junto con Charles Bergstresser, fundaron en 1882 Dow Jones & Company.

Estos promedios son revisados por editores del diario Wall Street Journal y sus componentes varían cuando hay una adquisición de por medio o cuando hay cambios drásticos en el desempeño corporativo de una empresa. Si ocurre algún evento que amerita el reemplazo de un componente, el índice al que este pertenece es revisado por entero. Por esta razón, múltiples cambios son realizados simultáneamente.

Aunque no hay regulaciones explícitas para la selección de los componentes de cada promedio, los títulos son añadidos únicamente si tienen una excelente reputación, demuestran crecimiento sustancial, son parte del interés de varios inversores y representan de forma adecuada el sector al cual pertenecen.

El Promedio Industrial Dow Jones (DJIA), a diferencia del Promedio de Transportes Dow Jones (DJTA) y el Promedio de Utilidades Dow Jones (DJUA), no sólo limita los títulos tradicionalmente definidos como industriales. Por el contrario, sus componentes pertenecen a todo tipo de industrias: servicios financieros, tecnología, minoristas, entretenimiento y bienes del consumidor.

El DJIA, es sin duda el más seguido y reconocido índice de acciones. Lo conforma un diversificado portafolio con los títulos de 30 empresas de valor fijo alto (*blue chips stocks*), es por esto que se lo considera como un medidor del desempeño del mercado en general.

Los promedios del Dow tienen la particularidad de que se calculan sobre la base del precio y no de la capitalización. Por ende, el peso de los componentes varía de acuerdo al precio de las acciones, a diferencia de otros índices que son afectados no sólo por el precio sino también por el número de acciones en circulación.

Cuando los promedios fueron creados, su valor era calculado simplemente por la suma del precio de todos los componentes de cada índice y dividiéndolo por el número de componentes del mismo. Luego, se inició la práctica de utilizar un divisor para suavizar los efectos de la división de acciones (*stock split*) y otras estrategias corporativas.<sup>34</sup>

### 1.2 Nasdaq

El Nasdaq es el primer mercado electrónico y el más grande del mundo. Sus negociaciones no están limitadas por un espacio físico, sino que éstas se realizan a través de un sofisticado sistema de computadoras y canales de telecomunicación.

El Nasdaq Stock Market cubre dos tipos de mercado: El Nasdaq National Market, en el cual cotizan acciones de empresas con una mayor capitalización, y el Nasdaq Smallcap Market, en el cual cotizan los títulos de pequeñas empresas con capacidad de crecimiento.

En la actualidad el Nasdaq enlista los títulos de alrededor de 4.100 empresas, desde

---

<sup>34</sup> [www.terra.com]

pequeñas firmas hasta gigantes de la talla de Microsoft. La mayoría de estos títulos pertenecen al sector tecnológico, telecomunicaciones, banca, minoristas y otras industrias en crecimiento.

Todas las firmas de negociación deben tener una certificación de la Comisión de Valores de los Estados Unidos (SEC) y deben estar registradas de acuerdo a las regulaciones del Nasdaq y el NASD.

### **1.3 IGPA**

El IGPA es un indicador de carácter patrimonial, que mide las variaciones de precios de la mayoría de las acciones inscritas en la Bolsa de Santiago y dicha medición se efectúa a través del Patrimonio Bursátil o Valor Bolsa de las diferentes sociedades que la componen, clasificadas en rubros y subrubros según su actividad. La cartera de este índice se re-evalúa el 30 de Diciembre de cada año, de acuerdo a la frecuencia en que se registran sus operaciones y a los volúmenes transados, y a la base de referencia del mismo que es 100 puntos a partir del 30 de Diciembre de 1980.

El criterio fundamental en la estructura del IGPA es medir variaciones de precios de las acciones inscritas oficialmente en Bolsa. Esta medición se efectúa a través del patrimonio bursátil (valor Bolsa) de las diferentes sociedades componentes, las cuales se han dividido en cinco rubros: Bancarias y Financieras; Agropecuarias y Forestales; Minerías; Industriales; Servicios varios. A su vez esos 5 rubros se dividen en doce subrubros.

$$IGPA = \frac{\sum PBR_k(1)}{\sum PBR_k(0)} * 100 \quad k= 1, \dots, 5 \quad (19)$$

Donde:

$PBR_k(1)$  = Patrimonio bursátil Rubro k en el momento 1.

$PBR_k(0)$  = Patrimonio bursátil Rubro k en la fecha base.

$$PBR_k = \sum_{i=1}^{40} (factor \ IGPA * precio \ cierre)_i^{35} \quad (20)$$

### **1.4 Tipo de Cambio**

El rol de las expectativas al determinar los tipos de cambio se desprende del hecho de que las monedas son activos financieros y de que una tasa cambiaria es simplemente el precio relativo entre dos de ellos. A diferencia de los precios de bienes y servicios, que no pueden ser almacenados por mucho tiempo, los precios de los activos financieros se van determinando por la voluntad de la gente de mantener dichos activos, que a su vez depende de las expectativas del valor de los mismos en el futuro.

El precio de una moneda dada dependerá de la propensión de los individuos a mantener sus inversiones bajo esta denominación. El tipo de cambio puede verse como el precio de equilibrio entre ofertas y demandas relativas de activos denominados en dos monedas.<sup>35</sup>

Estas variables las escogimos por la correlación e influencia que tienen sobre el tipo de cambio lo que puede ayudar a “explicar” sus movimientos futuros.

Nosotros configuramos modelos con estas variables y las fuimos alternando en las diferentes pruebas, con el fin de lograr una mejor predicción. Trataremos de identificar entre estas variables, cuáles son las que más están influyendo como también el nivel de rezago que mejores resultados entrega, para poder encontrar el modelo que mayor grado de acierto tiene en la predicción del cambio de signo en el precio del dólar.

## 2. Modelos

Para realizar nuestro estudio, utilizamos 254 datos de valores de cierre de diferentes variables, que están comprendidas entre el 3 de Junio de 1998 hasta el miércoles 30 de Abril de 2003. Los datos que utilizamos, son semanales, para lo cual elegimos el día miércoles de cada semana. Para evitar que algunas semanas quedaran sin datos producto de feriados, decidimos optar por la regla de tomar los martes en vez del miércoles en que no se tenga información. Si tampoco fuera posible esa alternativa, decidimos tomar el jueves.

La variable de salida la determinamos como la primera diferencia del valor del Dólar correspondiente al periodo “t” ( $\delta D_t$ ).

Las variables de entrada fueron tomadas rezagadas, ya que no se cuenta con la capacidad para predecir las variables utilizadas en el modelo. Podemos también agregar que la información de hoy está desfasada con respecto a la proyección, por lo que se utilizan modelos dinámicos con rezago ( $t-1$ ,  $t-2$ ,  $t-3$ ) en las variables independientes.

Las variables de entrada que escogimos son:

$$Dólar_{t-1}, Dólar_{t-2}, Dólar_{t-3}, Dow_{t-1}, Dow_{t-2}, IGFA_{t-1}, IGFA_{t-2} \quad (21)$$

El tamaño del grupo de entrenamiento lo vamos a determinar dada la conveniencia que nos reporta, entendiéndose como un tamaño óptimo aquel que permita darnos mejores resultados en cuanto al porcentaje de predicción de signo que deseamos obtener.

El tamaño del grupo extramuestral se va a determinar por la diferencia entre el total de datos menos aquellos utilizados para formar el conjunto de entrenamiento.

Dentro del programa Neuroshell, escogimos una arquitectura que nos facilite la predicción, por lo que escogimos para ello una red Ward. En ella usamos una función de escalamiento lineal (-1,1). Esto para que los datos se encuentren dentro de un rango más

<sup>35</sup> [Virreira 96e]

fácil de estudiar.

Dentro de esta arquitectura, tenemos que ocupar funciones que nos permitan transferir los datos de una capa a otra, por lo que se debe escoger una función determinada.

La capa de entrada tiene un slab y posee una neurona por cada variable de entrada. En nuestra capa oculta tenemos 3 slabs, con 3 neuronas cada uno, aquí vamos a ocupar una función de activación Gaussiana.

En nuestra capa de salida la función de activación va a ser una Tangente.

La salida neta de la capa oculta corresponde a la ecuación que sigue a continuación:

$$i_{pj}^k = f_j^k(\text{Neta}_{pj}^k) \quad , \text{ con } j=1, 2, \dots, k. \quad (22)$$

La ecuación que corresponde a la función neta que se incluyó en la expresión anterior es la siguiente:

$$\text{Neta}_{pj}^k = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1b} \\ w_{21} & w_{22} & \dots & w_{2b} \\ \vdots & \vdots & \dots & \vdots \\ w_{\lfloor \frac{j}{2} \rfloor 1} & w_{\lfloor \frac{j}{2} \rfloor 2} & \dots & w_{\lfloor \frac{j}{2} \rfloor b} \end{bmatrix}^k \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_b \end{bmatrix}_{(b \times 1)} + \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{\lfloor \frac{j}{2} \rfloor} \end{bmatrix}^k \left( \begin{bmatrix} \lfloor \frac{j}{2} \rfloor \\ 1 \end{bmatrix} \right) = \begin{bmatrix} \text{Neta}_{p1} \\ \text{Neta}_{p2} \\ \vdots \\ \text{Neta}_{p\lfloor \frac{j}{2} \rfloor} \end{bmatrix}^k \left( \begin{bmatrix} \lfloor \frac{j}{2} \rfloor \\ 1 \end{bmatrix} \right) \quad (23)$$

A continuación presentamos la salida neta de la capa de salida:

$$o_{pk} = f_k^o(\text{Neta}_{pk}^o) \quad (24)$$

Donde su función neta es:

$$\text{Neta}_{pk}^o = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1\lfloor \frac{k}{2} \rfloor} \\ w_{21} & w_{22} & \dots & w_{2\lfloor \frac{k}{2} \rfloor} \\ \vdots & \vdots & \dots & \vdots \\ w_{\lfloor \frac{k}{2} \rfloor 1} & w_{\lfloor \frac{k}{2} \rfloor 2} & \dots & w_{\lfloor \frac{k}{2} \rfloor \lfloor \frac{k}{2} \rfloor} \end{bmatrix}^o \times \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_{\lfloor \frac{k}{2} \rfloor} \end{bmatrix}^o \left( \begin{bmatrix} \lfloor \frac{k}{2} \rfloor \\ 1 \end{bmatrix} \right) + [\theta_1]_{(1 \times 1)}^o - [\text{Neta}_{pk}^o]_{(1 \times 1)}^o \quad (25)$$

### 3. Rolling

El funcionamiento Rolling consiste en agregar nuevos datos a los ya procesados por la red, pero se le entrega una mayor importancia a la información nueva. La explicación a esto se encuentra en el hecho de que los inversionistas van ajustando sus expectativas

(predicciones) dándole mayor importancia a la información más reciente y menos a la más antigua.

Se mantiene el tamaño del conjunto de entrenamiento “ $n$ ”, de manera que cada dato nuevo hace borrar el dato más antiguo conservando así siempre con “ $n$ ” el “tamaño de Rolling”. Este proceso se repite “ $m-1$ ” veces, haciendo que la red vuelva a calcular sus pesos de acuerdo a lo que ha aprendido. De esta forma, si consideramos que inicialmente el conjunto extramuestral tiene “ $m$ ” datos, al evaluar el funcionamiento de la red se consideró sólo la predicción del primer valor. Luego, el dato analizado sale del conjunto extramuestral (quedando con “ $m-1$ ” datos) y pasa a formar parte del conjunto de entrenamiento, pero manteniendo el tamaño de la muestra en “ $n$ ” datos por lo se agrega la información más reciente al conjunto extramuestral quedando nuevamente con “ $m$ ” datos, el cual ha mantenido su tamaño pero ha variado su composición. Luego se realizó una nueva iteración, lo que implicó reconstruir los pesos del modelo para cada una de las “ $m$ ” proyecciones que no han sido procesadas, permitiéndole a la red aprender del error cometido en la predicción y ajustar nuevamente los pesos estimados para los valores de entrada de las capas de salida y oculta,  $w_{jk}$  y  $yw_{jk}$ , respectivamente. Este proceso se repitió hasta el momento en que la última observación (correspondiente al período “ $t-1$ ”) es empleada para proyectar el valor que la variable de salida podría alcanzar en el momento “ $t$ ”, el cual representa el futuro inmediato.<sup>36</sup>

#### 4. Recursividad

El funcionamiento recursivo consiste en ir agregando información a la que ya se tiene para realizar una predicción. De esta manera si tengo un conjunto muestral (“Conjunto Recursivo inicial”) con “ $n$ ” datos ahora voy a ir agregando teniendo ahora “ $n+1$ ” observaciones. Se utiliza esta metodología ya que se considera que la información pasada va a influir en las variaciones futuras del tipo de cambio, ya que los agentes considerarán esta información al tomar decisiones de compra o venta. Al igual que cuando se utiliza la técnica Rolling este proceso se repite “ $m-1$ ” veces, haciendo que la red vuelva a calcular sus pesos de acuerdo a lo que ha aprendido

#### 5. Direccional Accuracy Test (DA)

Para evaluar la capacidad predictiva de cada modelo se utilizó el Direccional Accuracy Test (DA) de Pesaran & Timmermann<sup>37</sup>, para medir la significancia estadística de nuestros modelos de predicción. Lo que hace este test es comprobar si la proyección basada en el modelo y el retorno observado son independientes (es decir, si  $\hat{y}_{n+j}$  y  $y_{n+j}$  están independientemente distribuidos). Si se rechaza la hipótesis nula, significa que existe evidencia de que se cuenta con capacidad para predecir el comportamiento de la variable observada.

<sup>36</sup> [ParisiRoll 03a]

<sup>37</sup> [Pesaran 92]

## B. ANÁLISIS DE RESULTADOS

En nuestro trabajo fuimos analizando el porcentaje de predicción de signo (PPS) que se podía obtener de las variaciones del precio del tipo de cambio (\$/US\$) utilizando modelos de Redes Neuronales con las técnicas de Rolling y de Recursividad.

Para ello, quisimos evaluar diferentes variables relacionadas con la divisa. Éstas las escogimos por su influencia sobre el tipo de cambio y por la conveniencia en su utilización. Es por esto que fuimos alternando las variables con el fin de poder lograr el mayor porcentaje de predicción en cada caso.

Además de alternar las variables, quisimos probar la cantidad de datos óptimos para realizar la predicción. Para esto, probamos con tres opciones distintas, utilizando 20, 30 y 40 datos en el tamaño del Rolling (o "Conjunto Recursivo inicial", según corresponda) y 234, 224 y 214, respectivamente, en el conjunto extra muestral.

La forma funcional de los modelos predictivos analizados se presenta a continuación:

$$1. D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, Dow_{t-2}) \quad (26)$$

$$2. D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, Dow_{t-2}, IGPA_{t-1}) \quad (27)$$

$$3. D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, IGPA_{t-1}, IGPA_{t-2}) \quad (28)$$

$$4. D_t = f(D_{t-1}, D_{t-2}, IGPA_{t-1}, IGPA_{t-2}) \quad (29)$$

$$5. D_t = f(D_{t-1}, D_{t-2}, D_{t-3}, Dow_{t-1}, Dow_{t-2}) \quad (30)$$

$$6. D_t = f(D_{t-1}, D_{t-2}, D_{t-3}, Dow_{t-1}, IGPA_{t-1}) \quad (31)$$

Nuestro primer modelo lo realizamos con las variables: Dólar<sub>t-1</sub>, Dólar<sub>t-2</sub>, Dow<sub>t-1</sub> y Dow<sub>t-2</sub>, donde se ocuparon 20, 30 y 40 datos en el Rolling. Podemos ver que el modelo con funcionamiento Recursivo predice mejor que el modelo con la técnica Rolling para sus tres tamaños de Rolling (ver cuadro 1), lo que escapa a lo que los resultados normalmente muestran y a nuestras expectativas. A pesar de esto, podemos darnos cuenta que los resultados obtenidos logran un alto nivel de predicción para los tres grupos.

IV. MODELOS DE PREDICCIÓN DE REDES NEURONALES

Modelo funcional	PPS **	PPS **	Tamaño	Test DA	Test DA
	Recursivo	Rolling	Rolling	RECURSIVO	ROLLING
<b>1.</b> $D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, Dow_{t-2})$	69,23	67,52	20	6,047753*	5,210724*
	68,3	66,62	30	5,753247*	4,777431*
	67,23	66,35	40	5,396417*	4,919326*
<b>2.</b> $D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, Dow_{t-2}, IGPA_{t-1})$	67,1	66,05	20	4,713005*	5,677896*
	67,09	64,70	30	3,924133*	4,452159*
	62,14	66,22	40	3,520332*	5,668558*
<b>3.</b> $D_t = f(D_{t-1}, D_{t-2}, Dow_{t-1}, IGPA_{t-1}, IGPA_{t-2})$	65,81	70,08	20	4,671337*	5,159736*
	65,62	66,62	30	4,644316*	4,777431*
	64,43	67,28	40	4,263198*	5,341917*
<b>4.</b> $D_t = f(D_{t-1}, D_{t-2}, IGPA_{t-1}, IGPA_{t-2})$	68,37	66,05	20	5,691071*	5,633034*
	67,85	62,5	30	5,564715*	5,179403*
	66,82	66,82	40	5,203547*	5,163680*
<b>5.</b> $D_t = f(D_{t-1}, D_{t-2}, \Gamma_{t-2}, Dow_{t-1}, Dow_{t-2})$	67,09	70,5	20	5,15110*	5,190790*
	66,07	67,94	30	4,813341*	5,776437*
	64,95	64,95	40	4,435238*	4,406221*
<b>6.</b> $D_t = f(D_{t-1}, D_{t-2}, D_{t-3}, Dow_{t-1}, IGPA_{t-1})$	69,85	67,94	20	6,180131*	5,664619*
	69,61	66,83	30	6,210573*	4,070742*
	68,69	66,35	40	5,861438*	4,928607*

Cuadro 1: Resumen PPS, Tamaño Rolling, Test DA.

\*La capacidad predictiva es determinada con un nivel de confianza del 95%. Valor Z crítico = 1,96

\*\* Porcentaje de Predicción de Signo.

Además, nos podemos dar cuenta que los resultados del test DA nos indican que en todos los casos los resultados son significativos.

El hecho de que el modelo con funcionamiento Recursivo prediga mejor que el modelo Rolling quizás se podría explicar por las variables utilizadas y por el rezago de ellas. Podemos ver también que el tamaño de Rolling más adecuado para estas variables es de 20 datos.

En el segundo modelo ocupamos una variable más que en el primero, agregando el  $IGPA_{t-1}$ . Por lo tanto las variables utilizadas fueron:  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dow_{t-1}$ ;  $Dow_{t-2}$  e  $IGPA_{t-1}$ , con 20, 30 y 40 datos en el tamaño del Rolling. El resultado más relevante que

obtuvimos fue 69,65% de predicción de signo con un tamaño de Rolling de 20 datos.

Podemos observar que el modelo Rolling predice mejor en todos los casos que el modelo Recursivo. Esto concuerda con la teoría que dice que las personas valoran más la información más reciente que aquella que ocurrió hace ya un tiempo.

En el tercer modelo, ocupamos las mismas variables que en el segundo, la diferencia radica en el rezago que ellas presentan. Entonces las variables que encontramos son:

$Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dow_{t-1}$ ;  $IGPA_{t-1}$  e  $IGPA_{t-2}$ .

Se puede ver que el modelo Rolling tiene una mejor tasa de aciertos en dos casos (20 y 40 datos) y que logra igualar cuando el tamaño de Rolling es de 30 datos. Estos resultados también son consistentes con la teoría de que las personas le dan más importancia a la información más reciente. Podemos darnos cuenta que el mejor tamaño de Rolling para estas variables es de 20 datos, lo que se repite en los otros modelos. En este caso logramos sobrepasar la barrera del 70% de predicción.

En el cuarto modelo se utilizaron sólo dos variables, esta vez dejando de lado al Dow, quedando sólo con el IGPA y el Dólar con diferentes rezagos. Entre las variables encontramos:  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $IGPA_{t-1}$ ;  $IGPA_{t-2}$ . El resultado más relevante que obtuvimos fue 69,65% de predicción de signo con un tamaño de Rolling de 20 datos, al igual que en el segundo modelo.

En este caso, el modelo con funcionamiento Rolling solamente predice mejor cuando el tamaño de Rolling es de 20 datos e iguala al modelo con funcionamiento Recursivo en el caso en que el tamaño del Rolling es de 40 datos. En el segundo caso, cuando se utilizan 30 datos, la técnica Recursiva va a dar mejores resultados que la técnica de Rolling. En este caso es muy difícil saber lo que sucede, pero podríamos decir que las variables escogidas no serían las más adecuadas para poder determinar la técnica a ocupar, pero por los resultados generales sí podríamos ocupar estas variables para predecir el cambio del precio de la divisa, dado los altos porcentajes de predicción en los tres casos.

Podemos darnos cuenta que el test DA arroja resultados satisfactorios, señalando que los modelos evidencian una capacidad predictiva significativa en términos estadísticos.

En el quinto modelo ocupamos las mismas variables pero con los siguientes rezagos:  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dólar_{t-3}$ ;  $Dow_{t-1}$  y  $Dow_{t-2}$ .

La técnica Rolling nuevamente domina en cuanto a ser un mejor predictor, con el tamaño de Rolling de 20 datos. Lo más importante es que aquí se logra la mejor predicción de todas logrando un 70,51% de aciertos en el signo de la variación del tipo de cambio, mientras que el mismo modelo con funcionamiento Recursivo tenía un PPS de 67,09%.

En el sexto modelo volvemos a tomar las tres variables, aquí los rezagos también cambian, por lo que tenemos:  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dólar_{t-3}$ ;  $IPSA_{t-1}$  y  $Dow_{t-1}$ .

El modelo Recursivo predice mejor que el modelo Rolling en los tres casos, lo cual no coincide con la teoría que dice que el uso de la información más cercana tiene mayor

influencia que la información más alejada en el tiempo. Lo que sí se repite es que la mejor predicción se tiene con un tamaño de Rolling de 20 datos, lo que podría estar indicando un patrón.

Enfrentándonos a estos resultados, podemos destacar algunos puntos con respecto al comportamiento de nuestro estudio.

Vemos que en el quinto modelo se logra la mejor predicción con un 70,51% de predicción, seguido por el tercer modelo donde se logra un 70,08%. Si bien podríamos decir que el quinto modelo predice de mejor manera que el tercer modelo, en promedio el tercer modelo presenta mejores resultados y concuerda también con una estructura más adecuada, donde se integran las tres variables más relevantes (Dólar, Dow e IGPA), acentuando el rezago del dólar y del IGPA, como las variables más importantes para los inversionistas.

Otro factor a considerar es el tamaño del Rolling, vemos que en todos los modelos el tamaño óptimo es de 20 datos y se logran mejores resultados independientemente de las variables que se ocupen, como de los rezagos que estas tengan.

Vemos también que la técnica Rolling muestra una tendencia a ser mejor predictor que el funcionamiento Recursivo, esto se muestra ya que el porcentaje de predicción de signo de los modelos de Redes Neuronales Dinámicas para el caso de la técnica Rolling, supera al porcentaje de aciertos logrado por la técnica Recursiva.

Por otro lado, la teoría que indica que las personas son más sensibles a la información reciente, es un excelente argumento para apoyar la técnica de Rolling. Esto dado que su funcionamiento permite que los datos que se están tomando para realizar la predicción sean un número constante que se va actualizando agregando datos y restándole importancia a otros a medida que va transcurriendo el tiempo.

Un factor importante a considerar es el test Direccional Accuracy de Pesaran & Timmermann<sup>38</sup>, el cual nos indica si los resultados arrojados por el modelo son o no estadísticamente significativos. En nuestro estudio podemos ver que todos los resultados arrojaron valores significativos, lo cual es un signo importante para los inversionistas que tratan de minimizar el riesgo al momento de tomar decisiones.

---

<sup>38</sup> [Pesaran 92]



---

## V. CONCLUSIONES

El modelar mercados financieros a través de Redes Neuronales Dinámicas, requiere poseer una adecuada información de lo que se desea medir, como también de las variables que se van a utilizar, el número de datos que va a contener el tamaño del Rolling, la arquitectura entre otros. El poder determinar cada uno de estos elementos no fue fácil, pero finalmente se logró definir en nuestro trabajo seis modelos y tres variables, con diferentes desfases y/o rezagos para cada una de ellas y 254 datos divididos entre el tamaño del Rolling y un conjunto extramuestral, lo cual nos permitió poder comenzar a hacer el estudio.

Nuestro objetivo fue encontrar modelos de redes neuronales para predecir la variación del precio del tipo de cambio, comparando modelos que utilizaban un funcionamiento Rolling y otros una técnica Recursiva.

Los resultados que obtuvimos estuvieron de acuerdo a las expectativas que teníamos antes de comenzar nuestro estudio, lo que nos permite decir que la elección de nuestros datos, variables y modelos, fueron hechos de manera correcta, lo que además permite sacar buenas conclusiones.

Podemos decir que los resultados más importantes logrados en nuestro estudio superaron el 70% de predicción, esto es sin duda un resultado muy alentador para poder lograr una reducción de la incertidumbre al momento de tomar decisiones financieras. Los resultados corresponden al modelo que incluía las variables  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dow_{t-1}$ ;  $IGPA_{t-1}$  e  $IGPA_{t-2}$  y al modelo que contenía las variables  $Dólar_{t-1}$ ;  $Dólar_{t-2}$ ;  $Dólar_{t-3}$ ;  $Dow_{t-1}$  y  $Dow_{t-2}$ .

Con estos resultados sería fácil hablar de un modelo óptimo, pero para poder hacer una predicción no sólo se necesita contar con un buen modelo, sino que también se requiere de otros elementos que ayuden a lograr mejores resultados predictivos.

El poder determinar un buen modelo no es una tarea sencilla, también hay que analizar otros factores que influyen de manera significativa en el resultado. Dentro de estos factores podemos encontrar el tipo de técnica que debiera ser ocupada para poder predecir mejor el cambio del valor del dólar.

Entre las dos técnicas que analizamos, nos pudimos dar cuenta que la técnica de Rolling fue superior a la técnica Recursiva, dado el mayor porcentaje de predicción de signo que logró en los diferentes modelos donde se utilizó. Estas dos técnicas estuvieron soportadas por una arquitectura de red Ward, función mapeadora universal, la cual mostró resultados significativos en la utilización de los diferentes modelos.

Este resultado, sin duda, nos ayuda a reafirmar la teoría que dice que las personas, al tomar decisiones, lo hacen asignando una mayor importancia a la información más cercana que a aquella que se encuentra más alejada en el tiempo. Esto se ve reflejado en los resultados obtenidos en los diferentes modelos escogidos.

Es por esto, que la variabilidad de los pesos a través del tiempo es significativa y por lo tanto, el conjunto de entrenamiento no va a contener toda la información necesaria para entrenar la red, presentando evidencia en el sentido de que los precios responden a la formación de expectativas y estas expectativas cambian con el transcurso del tiempo ante el surgimiento de nuevos eventos.

Otro factor importante que hay que tener en cuenta, es el número de datos que se deben incluir en el Rolling (o en el "Conjunto Recursivo inicial", según corresponda). Dado los resultados obtenidos, nos pudimos dar cuenta que existe un patrón en los seis modelos que utilizamos. Este patrón nos indica que el tamaño óptimo de Rolling, debiera ser de 20 datos, esto equivaldría al 7,87% del conjunto. Estos resultados dan señales más claras a la hora de tomar decisiones financieras, ya que podemos establecer ciertos estándares que podrían ser aplicados en otros modelos.

Pero estos resultados no tendrían la validez necesaria si es que no son puestos a prueba para medir su capacidad predictiva. Nosotros utilizamos el test DA para medir este factor, y podemos decir que con un 95% de confianza, nuestros resultados son significativos en términos estadísticos.

En definitiva, los resultados han sido promisorios, dando paso a nuevas ideas y desafíos a realizar dentro de las finanzas. Probablemente el mayor aporte de este trabajo es haber planteado con nuevos modelos, la predicción de los cambios futuros en el precio del dólar y distinguiendo la eficacia del funcionamiento Rolling por sobre el Recursivo en modelos de Redes Neuronales Dinámicas, lo que sin duda creemos, va a incentivar el estudio de esta materia en las nuevas generaciones.

## VI. BIBLIOGRAFÍA

- [De Gregorio] De Gregorio, José, “Análisis empírico del tipo de cambio en Chile”  
Santiago: CEP, 1996, Cap. 2.
- [Del Brío 97a] Del Brío, Bonifacio Martín & Sanz Molina, Alfredo, “Redes Neuronales y Sistemas Borrosos”, Madrid, RAMA, 1997, p. 11, 12.
- [Espinoza 00] Espinoza Medina, Hernán Paul, “Predicción de Variables Financieras Utilizando Redes Neuronales”, Memoria de Ingeniero Civil Electricista, Universidad de Chile, 2000.
- [EspinozaMu 02a] Espinoza Muñoz, Christian, “Presentación Redes Neuronales: El Análisis Técnico en su expresión más avanzada”, Universidad de Chile, 2002.
- [Jiménez 96a] Jiménez Carvacho, Carlos, “Predicción de Acciones Utilizando Redes Neuronales”, Memoria de Ingeniero Civil Industrial, Universidad de Chile, 1996, p. 17.
- [Jiménez 96b] Jiménez Carvacho, Carlos, “Predicción de Acciones Utilizando Redes Neuronales”, Memoria de Ingeniero Civil Industrial, Universidad de Chile, 1996, p. 40, 41.
- [Jiménez 96c] Jiménez Carvacho, Carlos, “Predicción de Acciones Utilizando Redes Neuronales”, Memoria de Ingeniero Civil Industrial, Universidad de Chile, 1996, p. 18, 19, 20.
- [Larraín 90a] Larraín Sáenz, Roberto, “Análisis de las Variables Influyentes en la Predicción de Precios de Acciones”, Memoria de Ingeniero Industrial, Universidad de

Chile, 1990, p. 35.

- [Le Fort 97a] Le Fort, Guillermo, "Políticas, Condiciones y Perspectivas Macroeconómicas", Seminario "Proyecciones Económicas 1997-1998" organizado por Estrategia, Septiembre 1997, p. 1.
- [Le Fort 97b] Le Fort, Guillermo, "Políticas, Condiciones y Perspectivas Macroeconómicas", Seminario "Proyecciones Económicas 1997-1998" organizado por Estrategia, Septiembre 1997, p. 3.
- [Magendzo 96a] Magendzo W., Igal & Rojas R., Patricio & Vergara M., Rodrigo, "Análisis empírico del tipo de cambio en Chile", CEP, 1996, Capítulo 5.
- [Magendzo 96b] Magendzo W., Igal & Rojas R., Patricio & Vergara M., Rodrigo, "Análisis empírico del tipo de cambio en Chile", CEP, 1996, Capítulo 5.
- [Marshall 98] Marshall, Jorge, "Análisis de la Coyuntura Económica", Panel de ICU (Asociación de Ingenieros Comerciales de la Universidad de Chile), Septiembre 1998, p. 2.
- [Massad 98a] Massad, Carlos, "Políticas Macroeconómicas y Coyuntura", Taller de Coyuntura de la Universidad de Chile, Diciembre 1998, p. 1.
- [Massad 98b] Massad, Carlos, "Políticas Macroeconómicas y Coyuntura", Taller de Coyuntura de la Universidad de Chile, Diciembre 1998, p. 6.
- [O`Kean 94] O`Kean, José María, "Análisis del entorno económico de los negocios", Editorial McGraw-Hill, 1994, p. 109.
- [Parisi 02a] Parisi F., Antonino, "Evaluación de modelos de redes neuronales de predicción del signo de la variación del IPSA", Estudios de Administración, vol. 9. N°1, Universidad de Chile, 2002, p. 77.
- [Parisi 02b] Parisi F., Antonino, "Evaluación de modelos de redes neuronales de predicción del signo de la variación del IPSA", Estudios de Administración, vol. 9. N°1, Universidad de Chile, 2002, p. 78.
- [Parisi 02c] Parisi F., Antonino, "Evaluación de modelos de redes neuronales de predicción del signo de la variación del IPSA", Estudios de Administración, vol. 9. N°1, Universidad de Chile, 2002, p. 79.
- [Parisi 02d] Parisi, Antonino, "Evaluación de modelos de redes neuronales de predicción del signo de la variación del IPSA", Estudios de Administración, vol. 9, N° 1, Universidad de Chile, 2002, p. 71.
- [Parisi 02e] Parisi, Antonino, "Evaluación de modelos de redes neuronales de predicción del signo de la variación del IPSA", Estudios de Administración, vol. 9, N° 1, Universidad de Chile, 2002, p. 76.
- [ParisiRoll 03a] Parisi, Antonino, "Modelos de redes neuronales Rolling: tamaño muestral óptimo", Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2003.
- [ParisiOro 03a] Parisi, Antonino, "Modelos de redes neuronales Rolling versus Recursivo: Precio del Oro", Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2003.
- [Pesaran 92] Pesaran, M. H. & Timmermann A., "A Simple Non Parametric Test of Predictive Performance", Journal of business and economic statistics, 1992.

- 
- [Reyes 02a] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 7.
- [Reyes 02b] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 8, 9.
- [Reyes 02c] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 10.
- [Reyes 02d] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 11, 12.
- [Reyes 02e] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 13, 14.
- [Reyes 02f] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 15.
- [Reyes 02g] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 16.
- [Reyes 02h] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 17.
- [Reyes 02i] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 18.
- [Reyes 02j] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 32, 33.
- [Reyes 02k] Reyes, Daniel & Tapia, Víctor, “Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”, Facultad de Ciencias Económicas y Administrativas, Universidad de Chile, 2002, p. 42.
- [Rochette 94] Rochette García, Alain Emile, “Utilización del análisis técnico en los mercados financieros. Aplicación a la tasa de interés de largo plazo”. Memoria de Ingeniero Civil Industrial, Universidad de Chile, 1994.
- [Sanhueza 98] Sanhueza, Gonzalo, “Balanza Comercial y Política Cambiaria”, Jornadas de la Salmonicultura, Instituto Tecnológico del Salmón S.A., Septiembre de 1998, p. 7.
- [Virreira 96a] Virreira Zijderveld, Roberto, “Aplicación de Redes Neuronales a la Predicción y Transacción en Mercados Cambiarios”, Memoria Ingeniería Industrial de la Universidad Católica, 1996, p. 15.
- [Virreira 96b] Virreira Zijderveld, Roberto, “Aplicación de Redes Neuronales a la

Predicción y Transacción en Mercados Cambiarios”, Memoria Ingeniería Industrial de la Universidad Católica, 1996, p. 10.

[Virreira 96c] Virreira Zijderveld, Roberto, “Aplicación de Redes Neuronales a la Predicción y Transacción en Mercados Cambiarios”, Memoria Ingeniería Industrial de la Universidad Católica, 1996, p. 21.

[Virreira 96d] Virreira Zijderveld, Roberto, “Aplicación de Redes Neuronales a la Predicción y Transacción en Mercados Cambiarios”, Memoria Ingeniería Industrial de la Universidad Católica, 1996, p. 16.

[Virreira 96e] Virreira Zijderveld, Roberto, “Aplicación de Redes Neuronales a la Predicción y Transacción en Mercados Cambiarios”, Memoria Ingeniería Industrial de la Universidad Católica, 1996, p. 11.

[www.bolsadesantiago.com] <http://www.bolsadesantiago.com/>

[www.parisinet.com] <http://www.parisinet.com>

[www.terra.com] <http://www.terra.com/finanzas/articulo/html/fin517.htm>