



UNIVERSIDAD DE CHILE
FACULTAD DE ECONOMÍA Y NEGOCIOS
ESCUELA DE POSTGRADO, ECONOMÍA Y NEGOCIOS

“PREDICCIÓN DE SIGNO A TRES SEMANAS DE LA ACCIÓN CATERPILLAR CON REDES NEURONALES”

**TESIS PARA OPTAR AL GRADO DE
MAGISTER EN FINANZAS**

Profesor : Antonino Parisi F. Ph.D.

**Integrantes : Pedro Borquez Ramirez
Jorge Villanueva Ramos**

Santiago, 2006.-

ÍNDICE

1. INTRODUCCIÓN	3
2.- RESEÑA DE LA EMPRESA CATERPILLAR	8
2.1. Protagonista Estratégico De La Escena Mundial	8
2.2. Naturaleza de sus Operaciones	8
2.3. Factores De Riesgo de la Empresa	10
2.4. Hitos Históricos de la Compañía	11
3.-REDES NEURONALES.	14
3.1. Introducción a la Computación Neuronal	14
3.2. Historia de las Redes Neuronales	16
4.- METODOLOGÍA Y DATOS	20
4.1. Variables de Entrada – Series de Tiempo	20
4.2. Metodología Predictiva Utilizada	24
4.2.1. Redes Neuronales	24
4.2.2. Arquitectura de la Red Ward Usada en el Estudio	25
4.2.3. Arquitectura Multivariable Lineal	26
4.3. Construcción de los Modelos	28
4.4. Evaluación de la Capacidad Predictiva y Significancia Estadística.	29
4.5. Rentabilidad de Red Neuronal y Multivariable Lineal	29
5. ANÁLISIS DE LOS RESULTADOS OBTENIDOS	30
5.1. Metodología Red Neuronal Ward.	30
5.2. Modelo Multivariable Lineal	31
5.3. Convergencia de los Modelos	33
6. CONCLUSIONES	34
7. BIBLIOGRAFÍA Y REFERENCIAS	35
ANEXOS	



1.- 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 con el propósito de obtener mejores resultados frente a estos nuevos eventos. Dichos 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 el 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 están 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.

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 del tamaño muestral y la especificación del modelo. 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, nos aseguran un éxito relativamente mayor.

En otras palabras, el planteamiento del problema va a ser un elemento clave para lograr una mayor capacidad predictiva.

Además, hemos visto que la forma en que el modelo "estudia" o utiliza 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 modelo va agregando datos ha 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 predicción tenga mayor validez.

El estudio tiene por objetivo:

- ✓ Determinar el signo de la acción Caterpillar para tres semanas.
- ✓ Medir con test estadísticos los resultados obtenidos.
- ✓ Comparar la rentabilidad de estas estrategias de inversión con la de una estrategia pasiva, comprar mantener o “*buy and hold*” en los distintos escenarios (elegidos aleatoriamente) a fin de medir la significancia económica de los resultados y el cumplimiento o no de la hipótesis de mercados eficientes (Fama 1970), donde la eficiencia significa que el mercado refleja completa y correctamente toda la información relevante para la determinación de los precios de los activos.

De acuerdo con esto último, si los precios accionarios reflejaran instantáneamente toda la información relevante para la toma de decisiones, éstos cambiarían rápidamente ante la aparición de nueva información.

Luego, debido a que el surgimiento de nueva información es de carácter aleatorio, los cambios registrados por los precios accionarios también lo serían, invalidando el uso del análisis técnico ya que, en dicha situación, una regla de transacción basada únicamente en las series de precios históricos no puede ganar beneficios mayores que aquellos generados por una simple estrategia de comprar y mantener o “*buy and hold*” (Jensen & Benington, 1970) dado series anuales.

Por el contrario, varios estudios han concluido que existe evidencia significativa de que los precios accionarios no siguen un camino aleatorio y muestran que los retornos accionarios son predecibles en algún grado. Por ejemplo, Lo & MacKinlay (1988), empleando datos de mercados bursátiles desarrollados tales como EE.UU., Europa Occidental y Japón, registraron una correlación serial positiva entre los retornos semanales; Conrad & Kaul (1988, 1989) también encontraron evidencia de predictibilidad de los retornos en el corto plazo; DeBondt & Thaler (1985), e incluso el mismo Fama & French (1988), Poterba & Summers (1988) y Chopra, Lakonishok & Ritter (1992), hallaron una correlación serial negativa en los retornos de los





activos individuales y varios portafolios sobre intervalos de tres a diez años, es decir, en el largo plazo; Jegadeesh (1990), examinó la predictibilidad de los retornos mensuales sobre activos individuales y encontró una correlación serial negativa de primer orden altamente significativa para rezagos de dos meses y una correlación serial positiva para rezagos mayores; Blume, Easley & O'Hara (1994) presentaron un modelo que sugiere que existe una relación significativa entre los rezagos del volumen transado y los retornos actuales de los activos individuales.

Mientras hay consenso con respecto a que los retornos son predecibles, existe desacuerdo sobre las razones subyacentes a esta predictibilidad. Al momento de explicar la predictibilidad de las variaciones de los retornos accionarios se postulan dos argumentos alternativos:

- (a) los mercados son ineficientes y los precios de los activos se mueven alrededor de su valor fundamental, y
- (b) en los mercados eficientes y la predictibilidad de las variaciones puede ser explicada por un equilibrio en los retornos time-varying. Como Fama (1991) observa: “¿la predictibilidad de los retornos, refleja la desviación irracional del precio desde su valor fundamental, la variación racional a través del tiempo en los retornos esperados, o alguna combinación de estas dos situaciones?”.

Con relación al primer argumento, la hipótesis de “sobre-reacción del mercado accionario” enuncia que el mercado sobre-reacciona consistentemente a la nueva información y, en consecuencia, los precios accionarios se mueven alrededor de su valor fundamental debido a ondas de optimismo y pesimismo (ver, por ejemplo, DeBondt & Thaler [1985, 1987], Lehmann [1990], y Shefrin & Statman [1985]).

En uno de los trabajos más influyentes, DeBondt & Thaler (1985) presentaron evidencia de una reversión de los retornos económicamente importante sobre intervalos de largo plazo, donde activos que experimentaron un pobre desempeño durante un período de 3 a 5 años tienden a superar sustancialmente el desempeño de los activos ganadores del período previo durante los 3 a 5 años subsiguientes, lo cual es interpretado por los autores como una manifestación del comportamiento irracional de los inversores y que es denominado “sobre-reacción”.

Varios autores (por ejemplo, Chan [1988] y Ball & Kothari [1989]) han argumentado que la reversión de los retornos se debe principalmente a cambios sistemáticos en los retornos exigidos de equilibrio, que no son capturados por DeBondt & Thaler, siendo uno de los



principales los cambios significativos en el leverage de la compañía, el cual, junto con el riesgo del activo, afecta el beta de la firma.

Por su parte, Zarowin (1990) y otros han formulado que el mejor desempeño de los activos perdedores con respecto a los ganadores no se debe a una sobrereacción de los inversores, sino que es una manifestación del efecto tamaño y/o Enero en que, al final del período de ranking, los activos perdedores tienden a ser firmas de tamaño más pequeño que el de las ganadoras.

Continuando esta línea de investigación, Chopra, Lakonishok & Ritter (1992) encontraron un efecto de sobrereacción económicamente importante aún después de ajustar por tamaño y beta. No obstante, un estudio de Conrad & Kaul (1993) muestra que no hay evidencia de una sobrereacción del mercado y que el desempeño anormal de las estrategias de largo plazo evaluadas por diferentes investigadores (ver, por ejemplo, Ball & Kothari [1989]; Chan [1988]; Chopra, Lakonishok & Ritter [1992]; DeBondt & Thaler [1987]; y Zarowin [1990]) se debe a la combinación de una medida de desempeño sesgada y a un “efecto Enero”, el cual no está relacionado al desempeño pasado de los activos; Por su parte, Ferson & Harvey (1991) muestran que la predictibilidad de los retornos accionarios no se debe necesariamente a ineficiencias del mercado o a una sobrereacción de los inversionistas irracionales, sino que a la predictibilidad que presentan algunas variables agregadas que son parte del set de información que explica la rentabilidad de los activos.

Al margen de la controversia presentada en el párrafo anterior, al observar los modelos que examinan la importancia de la información contenida en las secuencias de precios, Brown & Jennings (1989) y Grundy & McNichols (1989) consideraron modelos de expectativas racionales en los cuales un único precio no revela la información subyacente, pero una secuencia de precios de activos sí lo hace, demostrando que el análisis técnico de patrones de precios puede ser valioso ya que facilita el aprendizaje de los traders.

Aún cuando varios estudios han mostrado que los retornos accionarios son predecibles en algún grado, en el caso de los activos individuales el grado de predictibilidad de los retornos es, generalmente, considerado económicamente no significativo. Al respecto, Brock, Lakonishok & LeBaron (1992) investigaron la predictibilidad de los retornos accionarios por medio del análisis técnico tradicional, cubriendo un período de 90 años sobre la base del índice Dow Jones Industry, y sus resultados proveen evidencia que soporta la idea de que el análisis técnico



produce retornos significativos en términos estadísticos y económicos, no obstante, los costos de transacción no fueron considerados¹. French y Roll (1986) encontraron una correlación serial negativa significativa en los retornos diarios, pero sugieren que es “pequeña en términos de magnitud absoluta” y que “es difícil medir su significancia económica”.

Por su parte, Lo & MacKinley (1988) consideraron rentabilidades semanales de activos individuales y encontraron que la correlación serial es estadística y económicamente no significativa, y sugieren que “el ruido idiosincrático... hace difícil detectar la presencia de componentes predecibles”. Sin embargo, estudios previos relativos a la predictibilidad de los retornos accionarios a partir de sus valores pasados sugieren que el análisis técnico podría ser útil. Parisi & Vasquez (2000) investigaron la capacidad predictiva de las medias móviles y de los canales más ampliamente usados, junto con la magnitud de los retornos que se derivan de estas técnicas para el mercado accionario de Chile, y encontraron que la media móvil de longitud variable (variable-length moving average) presenta una capacidad especial para evitar pérdidas. Además, los resultados del mencionado estudio validan las conclusiones de Gregoire (1985) y Urrutia (1994), relativas a la posibilidad de usar estrategias de inversión en Chile basadas en el análisis técnico².

En consecuencia, siendo la predicción de los retornos accionarios un hecho bien aceptado en la reciente literatura de las finanzas empíricas, las preguntas apuntan hacia cómo usar la información de una manera óptima para predecir la evolución futura de los precios de los activos y transar en los mercados.

¹ Si un indicador de análisis técnico desempeña su función, éste debe producir un retorno que exceda la suma de los costos de transacción más la rentabilidad negativa originada por sus señales falsas.

² Sin embargo, Parisi & Vasquez señalan que se debe considerar que los costos de transacción en el mercado de capitales chileno son elevados, lo cual podría limitar los resultados, acercándolos a aquellos obtenidos por Allen & Karjalainen (1999), donde el uso del análisis técnico produce retornos anormales desde el punto de vista estadístico, pero no económico.





2.- RESEÑA DE LA EMPRESA CATERPILLAR

2.1. Protagonista Estratégico De La Escena Mundial

Durante más de 80 años, Caterpillar Inc. ha estado construyendo infraestructuras en todo el mundo y, junto con su red de distribuidores, ha impulsado cambios positivos y sostenibles en todos los continentes. Con unas ventas de más de 36.340 millones de dólares en el año 2005, Caterpillar es una empresa líder en tecnología y el mayor fabricante del mundo de maquinaria de construcción y minería, de motores diesel y de gas natural y de turbinas de gas industriales.

Los 36.340 millones de ventas e ingresos de 2005 suponen una subida de 6.030 millones de dólares, o un 20%, con respecto a 2004. Esta subida es el resultado de un aumento de las ventas de 3.720 millones, 1.827 millones de dólares por subida de precios y 363 millones de aumento en ingresos por productos financieros. Aproximadamente la mitad de todas las ventas se hicieron a clientes de fuera de Estados Unidos, manteniendo la posición de Caterpillar como líder de suministro internacional y de exportación de Estados Unidos.

2.2. Naturaleza de sus Operaciones

Sus tres principales líneas de negocios son:

- 1.- Maquinarias: Es la principal línea de negocio la cual incluye el diseño, manufactura, marketing y venta para la construcción, minería y maquinaria forestal, camiones y cargadores, skidders, camiones de transportes, articulados, entre otros.
- 2.- Motores: La principal línea de negocios incluye motores para maquinaria Caterpillar, sistemas de generación eléctrica, vehículos de transporte, locomotoras, motores marinos, petroleros, industriales, agrícolas, y otras aplicaciones. Los motores van de 5 a 21.500 HP. Las turbinas están en un rango de 1.600 a 20.500 HP.
- 3.- Productos Financieros: Esta línea de negocio consiste fundamentalmente en la corporación de servicios financieros. Caterpillar provee financiamiento alternativo para sus clientes y representantes para maquinaria, motores y turbinas.



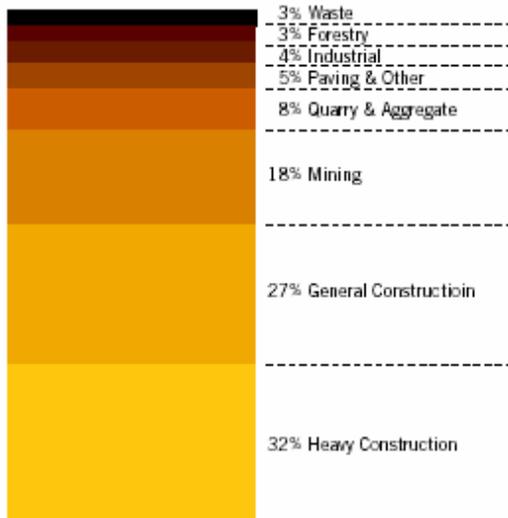


La distribución de sus ingresos está dado de acuerdo al cuadro siguiente:

Líneas de Negocios	Total 1er. Trimestre 2006	Porcentaje
Maquinaria	6.112	65%
Motores	2.631	28%
Productos Financieros	649	7%
Total	9.392	100%

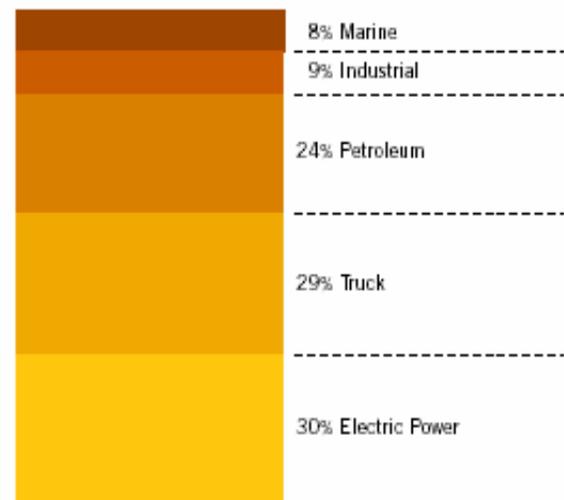
NEW MACHINERY DISTRIBUTION TO END USERS

(2005 dealer reported worldwide distribution of Caterpillar machinery by major end use)



NEW ENGINES DISTRIBUTION TO END USERS AND OEMS

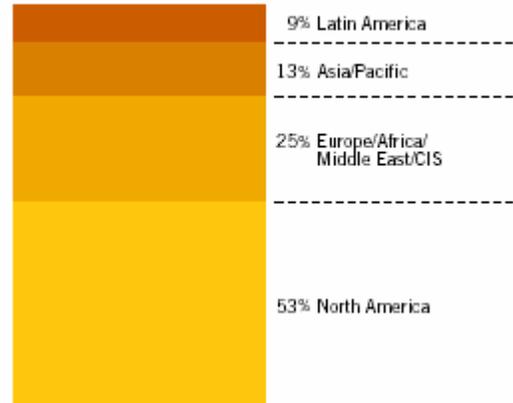
(2005 dealer reported worldwide distribution of Caterpillar engines by major end use)





2005 WORLDWIDE SALES AND REVENUES BY GEOGRAPHIC REGION

(2005 Caterpillar sales and revenues)



2.3. Factores de Riesgo de la Empresa

Los negocios de Caterpillar son fuertemente dependiente de la demanda de sus productos y una reducción importante en la demanda podría materializar efectos adversos en sus negocios, en su condición financiera, en su resultado operacional y flujo de caja.

La demanda por los productos Caterpillar y sus servicios, están influenciado por un número de factores incluidos entre otros:

- ✓ Condición económica general del mundo y el nivel de la minería, construcción y actividad manufacturera.
- ✓ Fluctuaciones en la demanda y precios de ciertos commodities.
- ✓ Fluctuaciones en el tipo de cambio y tasas de interés.
- ✓ Cambios e incerteza en la política monetaria fiscal de varios gobiernos y entidades reguladoras.
- ✓ La habilidad de Caterpillar para competir satisfactoriamente.
- ✓ Cambios en la política de precios de Caterpillar o de sus competidores.
- ✓ Cambios en la política económica y legislativa.
- ✓ Desastres naturales, guerras, embargos, actos terroristas y otros eventos catastróficos.



2.4. Hitos Históricos de la Compañía

1890. Benjamín Holt y Daniel Best experimentan con varias formas de tractores de vapor para su uso en granjas. Lo hacen separadamente, en compañías independientes.

1904. Primer tractor de cadenas de vapor de Holt.

1906. Primer tractor de cadenas de gas de Holt.

1915. Los tractores de cadenas "Caterpillar®" de Holt son usados por los Aliados en la Primera Guerra Mundial.

1925. Holt Manufacturing Company y C. L. Best Tractor Co. se unen para formar Caterpillar Tractor Co.

1931. Sale de la línea de producción el primer Tractor Sixty diesel en East Peoria (Illinois) con una nueva y eficaz fuente de propulsión para tractores de cadenas.

1940. La línea de productos Caterpillar incluye ahora motoniveladoras, hojas de nivelación, niveladores de elevación, levantadores de terrazas y grupos electrógenos para la generación de energía eléctrica.

1942. Los tractores de cadenas, motoniveladoras, grupos generadores y motores especiales de Caterpillar se utilizan en la aportación militar de Estados Unidos a la guerra para fabricar el tanque M4.

1950. Se establece Caterpillar Tractor Co. Ltd. en Gran Bretaña, primera compañía de operaciones internacionales creada para administrar la falta de cambio de moneda extranjera, tarifas, controlar la importación y servir mejor a clientes de todo el mundo.

1953. En 1931, la compañía crea un grupo de venta de motores independiente para vender motores diesel a otros fabricantes de equipo. Este grupo fue reemplazado en 1953 por una división independiente de ventas y mercadotecnia para servir mejor a una amplia variedad de clientes de motores. Las ventas de motores suponen aproximadamente una tercera parte de las ventas e ingresos totales de la compañía.



1963. Caterpillar y Mitsubishi Heavy Industries Ltd. forman una de las primeras empresas conjuntas de Japón para incluir propiedad parcial de Estados Unidos. Caterpillar Mitsubishi Ltd. comienza la producción en 1965, cambia su nombre a Shin Caterpillar Mitsubishi Ltd., y pasa a ser el fabricante número 2 de equipo para construcción y minería en Japón.

1981-83. La recesión mundial afecta a Caterpillar, costándole a la compañía el equivalente a \$1 millón al día y forzándole a reducir dramáticamente el número de empleados.

1983. Caterpillar Leasing Company se expande para ofrecer opciones de financiación de equipo a sus clientes de todo el mundo y cambia su nombre a Caterpillar Financial Services Corporation.

1985-presente. La línea de productos sigue diversificándose para satisfacer las necesidades de los clientes. Ahora se ofrecen más de 300 productos, más del doble de los que se ofrecían en 1981.

1986. Caterpillar Tractor Co. cambia su nombre a Caterpillar Inc., un reflejo más exacto de la creciente diversidad de la empresa.

1987. Se inicia un plan de modernización de fábricas de \$1.800 millones para mejorar el proceso de fabricación.

1990. La compañía descentraliza su estructura, reorganizándose en unidades comerciales para ofrecer mayor rendimiento de activos y satisfacción a los clientes.

1997. La compañía sigue expandiéndose, adquiriendo la empresa Perkins Engines, con sede en el Reino Unido. Con la incorporación de la alemana MaK Motoren el año anterior, Caterpillar se convierte en el líder mundial de fabricación de motores diesel.

1998. Aparece el camión de obras más grande del mundo -el 797- en los terrenos de pruebas de Cat en Arizona.



1999. Caterpillar desvela una nueva línea de equipo compacto para la construcción en la CONEXPO, la feria de construcción más grande del mundo, como respuesta a la cambiante necesidad de los clientes de equipo de construcción más pequeño y versátil.

2000. Caterpillar celebra su 75 aniversario.

2001. Caterpillar es la primera compañía en incorporar globalmente 6 Sigma y conseguir beneficios el primer año que exceden los costos de implementación.

2003. Caterpillar se convierte en el primer fabricante de motores del mundo ofreciendo una completa línea de motores diesel limpios para el año 2004 que cumplen todos los requisitos y certificaciones de la Environmental Protection Agency (EPA) de Estados Unidos. Se desarrolla la innovadora tecnología de control de emisiones de Caterpillar, conocida como Tecnología Avanzada de Reducción de Emisiones de Combustión (Advanced Combustion Emissions Reduction Technology - ACERT), para cumplir los requisitos de la EPA sin sacrificar rendimiento, fiabilidad o economía de consumo.



3.- REDES NEURONALES

3.1. Introducción a la Computación Neuronal

El hombre se ha caracterizado siempre por su búsqueda constante de nuevas vías para mejorar sus condiciones de vida. Estos esfuerzos le han servido para reducir el trabajo en aquellas operaciones en las que la fuerza juega un papel primordial. Los progresos obtenidos han permitido dirigir estos esfuerzos a otros campos, como por ejemplo, a la construcción de máquinas calculadoras que ayuden a resolver de forma automática y rápida determinadas operaciones que resultan tediosas cuando se realizan a mano.

Charles Babbage trató de construir una máquina capaz de resolver problemas matemáticos. Posteriormente otros tantos intentaron construir máquinas similares, pero no fue hasta la Segunda Guerra Mundial, cuando ya se disponía de instrumentos electrónicos, que se empezaron a recoger los primeros frutos. En 1946 se construyó la primera computadora electrónica, ENIAC. Desde entonces los desarrollos en este campo han tenido un auge espectacular.

Estas máquinas permiten implementar fácilmente algoritmos para resolver multitud de problemas que antes resultaban engorrosos de resolver. Sin embargo, se observa una limitación importante: ¿qué ocurre cuando el problema que se quiere resolver no admite un tratamiento algorítmico, como es el caso, por ejemplo, de la clasificación de objetos por rasgos comunes? Este ejemplo demuestra que la construcción de nuevas máquinas más versátiles requiere un enfoque del problema desde otro punto de vista. Los desarrollos actuales de los científicos se dirigen al estudio de las capacidades humanas como una fuente de nuevas ideas para el diseño de las nuevas máquinas. Así, la inteligencia artificial es un intento por descubrir y describir aspectos de la inteligencia humana que pueden ser simulados mediante máquinas. Esta disciplina se ha desarrollado fuertemente en los últimos años teniendo aplicación en algunos campos como visión artificial, demostración de teoremas, procesamiento de información expresada mediante lenguajes humanos... etc.

Las redes neuronales son más que otra forma de emular ciertas características propias de los humanos, como la capacidad de memorizar y de asociar hechos. Si se examinan con atención aquellos problemas que no pueden expresarse a través de un algoritmo, se observará que todos ellos tienen una característica en común: la experiencia. El hombre es capaz de resolver



estas situaciones acudiendo a la experiencia acumulada. Así, parece claro que una forma de aproximarse al problema consista en la construcción de sistemas que sean capaces de reproducir esta característica humana.

En definitiva, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el ejemplo más perfecto del que disponemos para un sistema que es capaz de adquirir conocimiento a través de la experiencia. Una red neuronal es "un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: *la neurona*".

Todos los procesos del cuerpo humano se relacionan en alguna u otra forma con la inactividad de estas neuronas. Las mismas son un componente relativamente simple del ser humano, pero cuando millares de ellas se conectan en forma conjunta se hacen muy poderosas.

También, es bien conocido que los humanos son capaces de aprender. Aprendizaje significa que aquellos problemas que inicialmente no pueden resolverse, pueden ser resueltos después de obtener más información acerca del problema.

Por lo tanto, las Redes Neuronales:

Consisten de unidades de procesamiento que intercambian datos o información.

Se utilizan para reconocer patrones, incluyendo imágenes, manuscritos y secuencias de tiempo, tendencias financieras.

Tienen capacidad de aprender y mejorar su funcionamiento.

Una primera clasificación de los modelos de redes neuronales podría ser, atendiendo a su similitud con la realidad biológica:

- 1) El modelo de tipo biológico. Este comprende las redes que tratan de simular los sistemas neuronales biológicos, así como las funciones auditivas o algunas funciones básicas de la visión.
- 2) El modelo dirigido a aplicación. Este modelo no tiene por qué guardar similitud con los sistemas biológicos. Su arquitectura está fuertemente ligada a las necesidades de las aplicaciones para la que es diseñada.



3.2. Historia de las Redes Neuronales

Conseguir diseñar y construir máquinas capaces de realizar procesos con cierta inteligencia ha sido uno de los principales objetivos de los científicos a lo largo de la historia. De los intentos realizados en este sentido se han llegado a definir las líneas fundamentales para la obtención de máquinas inteligentes: En un principio los esfuerzos estuvieron dirigidos a la obtención de **autómatas**, en el sentido de máquinas que realizaran, con más o menos éxito, alguna función típica de los seres humanos. Hoy en día se continúa estudiando en ésta misma línea, con resultados sorprendentes, existen maneras de realizar procesos similares a los inteligentes y que podemos encuadrar dentro de la llamada **Inteligencia Artificial (IA)**.

A pesar de disponer de herramientas y lenguajes de programación diseñados expresamente para el desarrollo de máquinas inteligentes, existe un enorme problema que limita los resultados que se pueden obtener: estas máquinas se implementan sobre computadoras basadas en la filosofía de Von Neumann, y que se apoyan en una descripción secuencial del proceso de tratamiento de la información. Si bien el desarrollo de estas computadoras es espectacular, no deja de seguir la línea antes expuesta: una máquina que es capaz de realizar tareas mecánicas de forma increíblemente rápida, como por ejemplo cálculo, ordenación o control, pero incapaz de obtener resultados aceptables cuando se trata de tareas como reconocimiento de formas, voz, etc.

La otra línea de la investigación ha tratado de aplicar principios físicos que rigen en la naturaleza para obtener máquinas que realicen trabajos pesados en nuestro lugar. De igual manera se puede pensar respecto a la forma y capacidad de razonamiento humano; se puede intentar obtener máquinas con esta capacidad basadas en el mismo principio de funcionamiento.

No se trata de construir máquinas que compitan con los seres humanos, sino que realicen ciertas tareas de rango intelectual con que ayudarle, principio básico de la Inteligencia Artificial. Las primeras explicaciones teóricas sobre el cerebro y el pensamiento ya fueron dadas ya por Platón (427-347 a.C.) y Aristóteles (348-422 a.C.). Las mismas ideas también las mantuvieron Descartes (1569-1650) y los filósofos empiristas del siglo XVIII.

La clase de las llamadas *máquinas cibernéticas*, a la cual la computación neuronal pertenece, tiene más historia de la que se cree: Herón (100 a.C) construyó un autómata hidráulico.



1936 - Alan Turing. Fue el primero en estudiar el cerebro como una forma de ver el mundo de la computación. Sin embargo, los primeros teóricos que concibieron los fundamentos de la computación neuronal fueron Warren McCulloch, un neurofisiólogo, y Walter Pitts, un matemático, quienes, en 1943, lanzaron una teoría acerca de la forma de trabajar de las neuronas (Un Cálculo Lógico de la Inminente Idea de la Actividad Nerviosa - Boletín de Matemática Biofísica 5: 115-133). Ellos modelaron una red neuronal simple mediante circuitos eléctricos.

1949 - Donald Hebb. Escribió un importante libro: *La organización del comportamiento*, en el que se establece una conexión entre psicología y fisiología. Fue el primero en explicar los procesos del aprendizaje (que es el elemento básico de la inteligencia humana) desde un punto de vista psicológico, desarrollando una regla de como el aprendizaje ocurría. Aun hoy, este es el fundamento de la mayoría de las funciones de aprendizaje que pueden hallarse en una red neuronal. Su idea fue que el aprendizaje ocurría cuando ciertos cambios en una neurona eran activados. También intentó encontrar semejanzas entre el aprendizaje y la actividad nerviosa. Los trabajos de Hebb formaron las bases de la Teoría de las Redes Neuronales.

1950 - Karl Lashley. En sus series de ensayos, encontró que la información no era almacenada en forma centralizada en el cerebro sino que era distribuida encima de él.

1956 - Congreso de Dartmouth. Este Congreso frecuentemente se menciona para indicar el nacimiento de la inteligencia artificial.

1957 - Frank Rosenblatt. Comenzó el desarrollo del Perceptrón. Esta es la red neuronal más antigua; utilizándose hoy en día para aplicación como reconocedor de patrones. Este modelo era capaz de generalizar, es decir, después de haber aprendido una serie de patrones podía reconocer otros similares, aunque no se le hubiesen presentado anteriormente. Sin embargo, tenía una serie de limitaciones, por ejemplo, su incapacidad para resolver el problema de la función OR-exclusiva y, en general, era incapaz de clasificar clases no separables linealmente. En 1959, escribió el libro Principios de Neurodinámica, en el que confirmó que, bajo ciertas condiciones, el aprendizaje del Perceptrón convergía hacia un estado finito (Teorema de Convergencia del Perceptrón).

1960 - Bernard Widrow/Marcial Hoff. Desarrollaron el modelo Adaline (ADAPTative LINear Elements). Esta fue la primera red neuronal aplicada a un problema real (filtros adaptativos para





eliminar ecos en las líneas telefónicas) que se ha utilizado comercialmente durante varias décadas.

1961 - Karl Steinbeck: Die Lernmatrix. Red neuronal para simples realizaciones técnicas (memoria asociativa).

1967 - Stephen Grossberg. A partir de sus conocimientos fisiológicos, ha escrito numerosos libros y desarrollado modelo de redes neuronales. Realizó una red: **Avalancha**, que consistía en elementos discretos con actividad que varía en el tiempo que satisface ecuaciones diferenciales continuas, para resolver actividades como reconocimiento continuo de habla y aprendizaje de los brazos de un robot.

1969 - Marvin Minsky/Seymour Papert. En este año surgieron críticas que frenaron, hasta 1982, el crecimiento que estaban experimentando las investigaciones sobre redes neuronales. Minsky y Papera, del Instituto Tecnológico de Massachusetts (MIT), publicaron un libro *Perceptrons. Probaron* (matemáticamente) que el Perceptrón no era capaz de resolver problemas relativamente fáciles, tales como el aprendizaje de una función no-lineal. Esto demostró que el Perceptrón era muy débil, dado que las funciones no-lineales son extensamente empleadas en computación y en los problemas del mundo real. A pesar del libro, algunos investigadores continuaron su trabajo. Tal fue el caso de **James Anderson**, que desarrolló un modelo lineal, llamado **Asociador Lineal**, que consistía en unos elementos integradores lineales (neuronas) que sumaban sus entradas. Este modelo se basa en el principio de que las conexiones entre neuronas son reforzadas cada vez que son activadas. Anderson diseñó una potente extensión del Asociador Lineal, llamada **Brain State in a Box (BSB)**.

1974 - Paul Werbos. Desarrolló la idea básica del algoritmo de aprendizaje de *propagación hacia atrás* (backpropagation); cuyo significado quedó definitivamente aclarado en 1985.

1977 - Stephen Grossberg. Teoría de Resonancia Adaptada (TRA). La Teoría de Resonancia Adaptada es una arquitectura de red que se diferencia de todas las demás previamente inventadas. La misma simula otras habilidades del cerebro: memoria a largo y corto plazo.

1977 - Teuvo Kohonen. Ingeniero electrónico de la Universidad de Helsinki, desarrolló un modelo similar al de Anderson, pero independientemente.



1980 - Kunihiko Fukushima. Desarrolló un modelo neuronal para el reconocimiento de patrones visuales.

1985 - John Hopfield. Provocó el renacimiento de las redes neuronales con su libro: "Computación neuronal de decisiones en problemas de optimización."

1986 - David Rumelhart/G. Hinton. Redescubrieron el algoritmo de aprendizaje de propagación hacia atrás (backpropagation). A partir de 1986, el panorama fue alentador con respecto a las investigaciones y el desarrollo de las redes neuronales. En la actualidad, son numerosos los trabajos que se realizan y publican cada año, las aplicaciones nuevas que surgen (sobre todo en el área de control) y las empresas que lanzan al mercado productos nuevos, tanto hardware como software (sobre todo para simulación).



4.- METODOLOGÍA Y DATOS

Esta investigación se centra en el hecho de que a través de la aplicación de redes neuronales con funcionamiento Aleatorio se puede lograr modelos que cumpla con las propiedades:

1. que cumplan con el nuevo paradigma de la supuesta no normalidad en los mercados financieros o que al menos no dependan explícitamente de ella.
2. que sean sistemas cuyas variables endógenas y exógenas puedan presentar relaciones no lineales entre ellas.
3. y que se adapten a la nueva información en la medida que ésta está disponible y por lo tanto, reflejen el cambio de las expectativas de los agentes presentes en los mercados financieros.

Así, se pretende reflejar que los precios responden a la formación de expectativas de los agentes del mercado, y éstas expectativas cambian con el transcurso del tiempo ante el surgimiento de nuevos eventos, cuestión que no es tomada en consideración por el funcionamiento estándar de la red neuronal. Para ello, la red es constantemente reentrenada y probada ante datos que nunca ha visto.

4.1. Variables de Entrada – Series de Tiempo

Para la realización del estudio, utilizamos 232 observaciones cada 3 semanas del cambio en el precio de la acción de Cat correspondientes al período entre el 8 de enero de 1993 y 21 de abril del 2006.

A través de la metodología del análisis fundamental determinamos las variables de entrada que de alguna forma se correlacionaba con la variable en estudio. Este análisis arrojó como una de las variables de entrada los competidores, los clientes, la evolución del Mercado de USA y otros como del Asia Pacífico y Sudamérica.



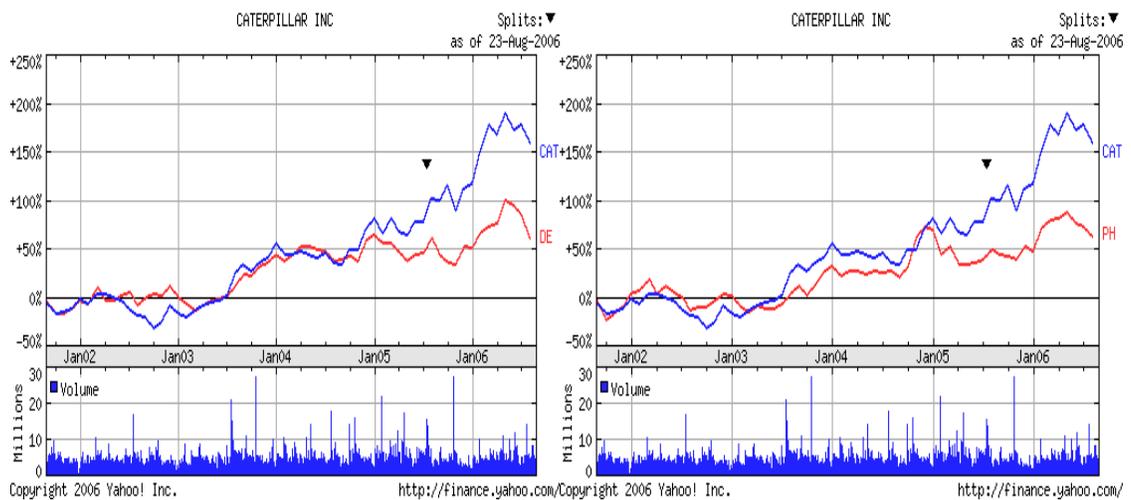
Las variables y sus gráficos de correlación se muestran en los cuadros siguientes

BHP	PD
BHP Billiton Limited	Phelps Dodge
Minera	Minera
Cientes	

VOLV	DE	PH
Volvo	Deere & Co	Parker Hannifin CP
Competidor Construcción Maquinarias	Competidor Construcción Maquinarias	Competidor Construcción Maquinarias
Competidores		

^N225	^MXX	^DJI	^IXIC	^GSPC	^TNX	CH
Nikkei 225	IPC de México	Dow Jones	Nasdaq	S&P 500	10-year treasury note	Chile Fund Inc
Indice mercado	Indice mercado	Indice mercado	Indice mercado	Indice mercado		Fondo Inversiones Chile
Mercados						

Cuadro N° 1



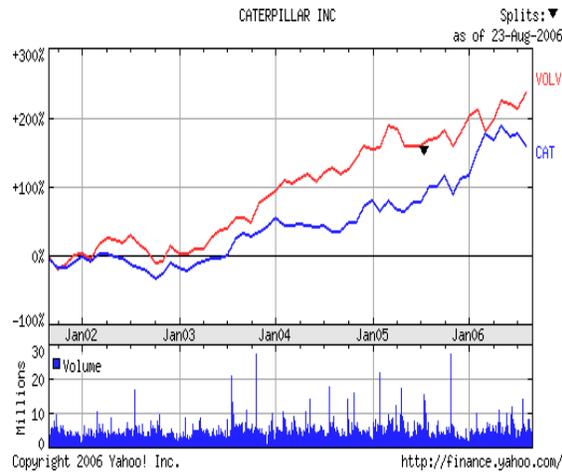


Grafico N° 1
Precio Acción Caterpillar versus Competidores

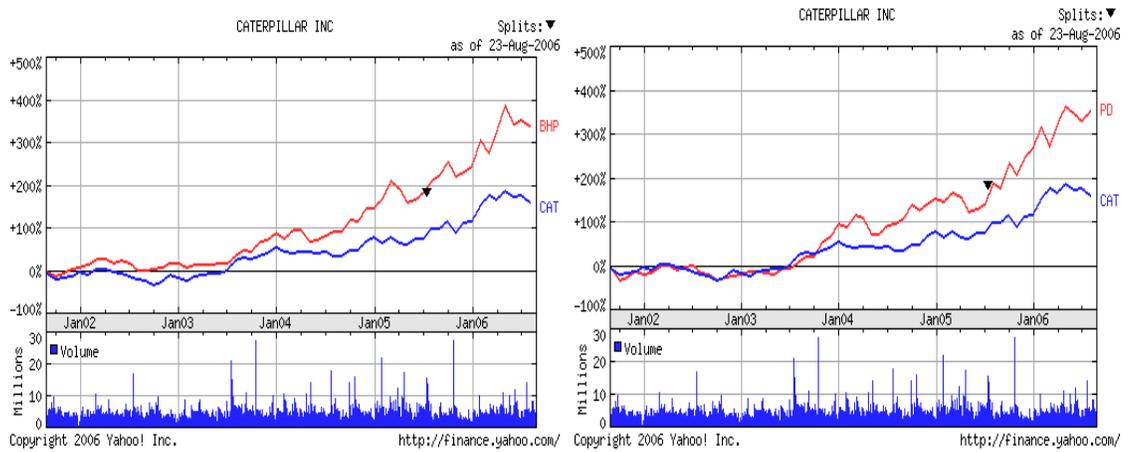


Gráfico N° 2
Precio Acción Caterpillar versus Clientes

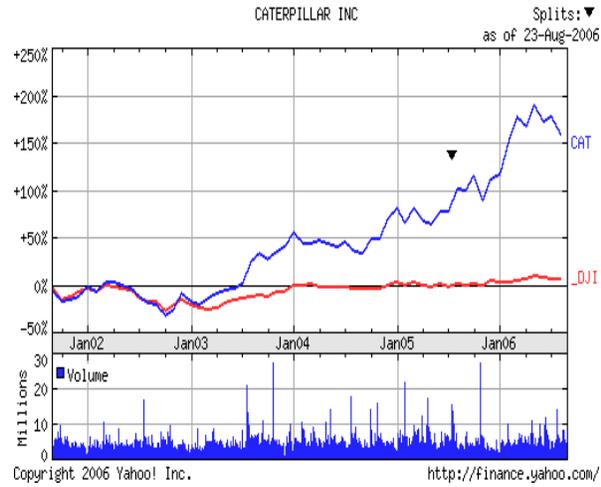


Gráfico N° 3
Precio Acción Caterpillar versus Dow Jones

Los gráficos 1, 2 y 3 muestran las variables de entradas expresadas en precio y su correlación respecto del precio de la acción Caterpillar.



4.2. Metodología Predictiva Utilizada

4.2.1. Redes Neuronales

En base a los trabajos efectuados por Parisi y Parisi (2003), Gonzales y Jimenez (2003) y Diaz (2005) donde probaron diferentes arquitecturas para las Redes Neuronales Artificiales y los estudios de Parisi, Parisi y Cornejo (2003) donde investigan modelos multivariados ARIMAX dinámicos a partir de arquitecturas basadas en algoritmos genéticos, la presente investigación se desarrolla bajo el mismo esquema de trabajo.

Funcionamiento de una neurona artificial; aplica una función (f) bajo un criterio de aplicación (α), a unas datos de entrada (x_i) que tienen cada uno de ellos una influencia (β_i), produciendo una salida (y_i).

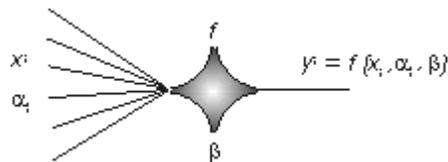


Figura N° 1

Para el caso de las redes neuronales se utilizo la Red Ward, la cual es capaz de detectar diferentes formas y patrones lineales y no lineales en los datos, a través de diferentes grupos de neuronas y diferentes funciones de activación en sus capas ocultas. Esta red posee 2 o mas capas ocultas cada una con una función de activación distinta. Sea f la función teórica, dinámica y desconocida que explica la tendencia alcista o bajista del precio de la acción de Caterpillar, la cual cuenta con un vector de variables explicativas, definidas anteriormente, relacionadas a la variable dependiente $\Delta(\text{Cat}) = g(X)$. Así, f puede ser aproximada por una red neuronal multicapa.

La forma funcional general de los modelos es expresada en la siguiente ecuación:

$$\Delta \text{Cat}_t = g(\Delta G_{t-i}, \Delta H_{t-i}, \Delta J_{t-i}, \Delta K_{t-i}, \Delta Q_{t-i})$$



Donde,

G, H, J, K y Q corresponden a las variables explicativas definidas en los modelos.

Las variables de entrada corresponden a la diferencia del precio o indicador en porcentaje, rezagada desde 1 a 2 períodos.

Los modelos consideran como variable de salida la predicción del signo de la acción Cat en un futuro próximo de tres semanas, el cual representa nuestro futuro inmediato (es decir, one step ahead).

4.2.2. Arquitectura de la Red Ward Usada en el Estudio

La red neuronal Ward puede ser escrita como una función de n (numero de neuronas de la capa oculta) variables explicativas x_i ($i = 1, 2, 3, \dots, 7$) definidas en X , con funciones de activación Tangente Hiperbólica, Gaussiana, Gaussiana Complementaria y SalidaHipertangente. En la figura N° 2 se muestra gráficamente la arquitectura de la red Ward, la cual básicamente consiste en 1 slab de entrada con 4 neuronas, slab 2 y 3 ocultas con 6 neuronas cada una y un 4 slab con 1 neurona de salida.

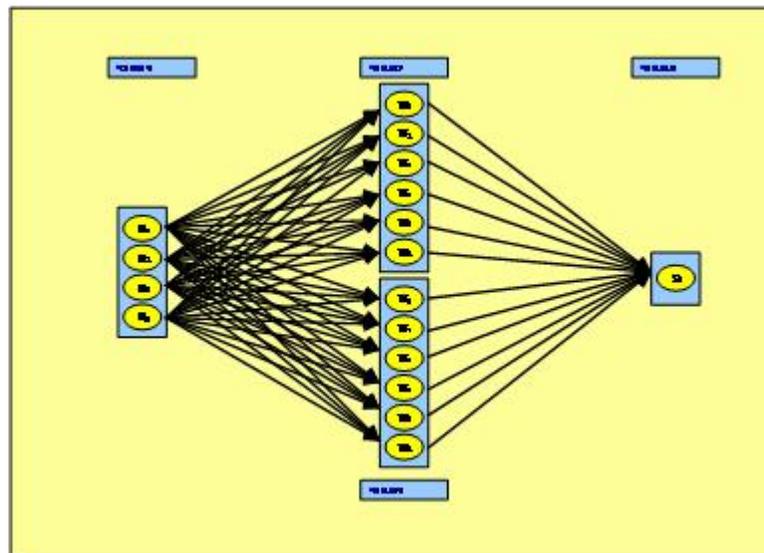


Figura N° 2

**Red neuronal Ward usada en el la investigación:**

	Slab 1	Slab 2	Slab 3	Slab 4	
Funciones	Tangenth	Gaussiana	ComGaussiana	Tanh	Total
Neuronas	4	6	6	1	17

Cuadro N° 3

4.2.3. Arquitectura Lineal Multivariable

En 1970, los estadísticos británicos Box y Jenkins desarrollaron una metodología destinada a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo desempeña un rol fundamental. Parte importante de esta nueva metodología se orientaba a eliminar la tediosa tarea de especificar un modelo, dejando que sea los propios datos temporales de la variable a estudiar indiquen la estructura probabilística subyacente, contraponiéndose a la forma estándar o tradicional de identificación y especificación de modelos. Estos nuevos modelos desarrollados se caracterizan porque una variable explicativa utiliza exclusivamente una variable exógena que es su propio pasado.

Dentro de los modelos uní variantes mas conocidos son los modelos ARIMA. Los modelos ARIMA o Auto regresivos Integrados de Medias Móviles son modelos econométricos en los cuales uno se aproxima a una nube de puntos a través de una regresión lineal, ajustando los ponderadores (W_i) de las variables explicativas para obtener el menor error cuadrático medio posible, en este caso, la regresión de las variables explicativas son los rezagos de la misma variable, los errores de predicción de los periodos anteriores y variables exógenas.

Se define a un modelo como auto regresivo si la variable endógena de un periodo t es explicada por las observaciones de ella misma correspondientes a periodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos indica que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) mas un termino de error (innovación).

La expresión genérica de un modelo auto regresivo, $AR(p)$ seria la siguiente:



$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon$$



4.3. Construcción de los Modelos

Con los antecedentes anteriores se construyó un cuadro con los Modelos para el estudio. La elección de variables fue aleatoria, utilizando hasta dos rezagos en cada una de ellas. El cuadro siguiente muestra las distintas combinaciones arrojando 12 modelos incluidos en el estudio.

Modelos		CAT 1	CAT 2	CAT 3	CAT 4	CAT 5	CAT 6	CAT 7	CAT 8	CAT 9	CAT 10	CAT 11	CAT 12
Minera	BHP(-1)	1		1						1			
	BHP(-2)		1								1		1
	PD (-1)				1	1			1				
	PD (-2)						1					1	
Competidores	VOLV(-1)	1						1					
	VOLV(-2)		1						1				
	PH(-1)				1						1		
	PH(-2)								1				1
	DE(-1)			1						1			
	DE(-2)					1		1					
Indices	DJI(-1)	1		1									
	DJI(-2)		1				1						
	IXIC(-1)								1			1	
	IXIC(-2)									1			
	MXX(-1)						1						
	MXX(-2)				1						1		
	N225(-1)							1					
	N225(-2)					1						1	
	GSPC(-1)							1					1
GSPC(-2)												1	
Letras y Fondos	TNX(-1)	1		1	1								
	TNX(-2)		1								1		
	CH(-1)						1						1
	CH(-2)					1				1			

Cuadro N° 4

De esta forma el **modelo CAT 1**, queda constituido con la variable de entrada número uno correspondiente a la primera diferencia porcentual del precio de la acción del **cliente** BHP con un rezago de la serie trisemanal. La variable de entrada número dos corresponde a la primera diferencia porcentual de precio de la acción del **competidor** VOLV con un rezago de la serie trisemanal. La variable de entrada número tres corresponde a la primera diferencia porcentual de precio de la acción del **Indice** DJI con un rezago de la serie trisemanal. La variable de entrada número cuatro corresponde a la primera diferencia porcentual de precio de los **Treasury Bond** TNX con un rezago de la serie trisemanal.

Los siguientes modelos que muestra el cuadro se construyeron de la misma forma.



4.4. Evaluación De La Capacidad Predictiva Y Significancia Estadística.

El proceso de evaluación empírica se hizo sobre la base de los datos del conjunto extramuestral. En ésta etapa, el desempeño relativo de los modelos fue medido por el número de predicciones correctas (hits) del signo de la variación del índice. Luego, al igual que Kanas (2001), se aplicó el directional accuracy test (DA T) de Pesaran & Timmermann (1992), con el objeto de medir la significancia estadística de la capacidad predictiva de los modelos de proyección. Pesaran & Timmermann (1992) mostraron que, bajo la hipótesis nula de que y_{n+i} y Y_{n+i} están independientemente distribuidos, dicho test sigue una distribución normal estándar.

En el cuadro de resultados, se muestran los resultados del Test obtenidos para cada modelo.

4.5. Rentabilidad de Red Neuronal y Multivariable Lineal

A ambos métodos, se agregó, el cálculo de la rentabilidad obtenida de seguir las recomendaciones de compra cuando el valor proyectado para el futuro cercano indicaba un alza en el precio, y una recomendación de venta en el caso contrario. El cálculo de las rentabilidades de las técnicas, consideró un monto de inversión inicial de US\$10.000 de manera referencial, valorándose la estrategia como el valor total de la cartera, es decir, la suma del inventario de la acción, más el dinero en efectivo con el que se contaba. Las señales de compra fueron ejecutadas en la medida que la cartera contaba con dinero en efectivo al momento de la señal, y las recomendaciones de ventas fueron ejecutadas en la medida que se contaba con las acciones. No se consideraron ventas cortas, costos de transacción, ni de almacenaje.

En los Cuadros, 5 y 6 se muestran las distribuciones encontradas para el PPS extramuestral de la mejor Red Ward y el mejor modelo Lineal Multivariado, junto con las distribuciones de las rentabilidades respectivas a cada técnica.



5. ANÁLISIS DE LOS RESULTADOS OBTENIDOS

5.1. Metodología Red Neuronal Ward.

Respecto de esta metodología, que consistió básicamente en ejecutar 12 modelos con distintas variables de entrada (ver punto 4.3.), y mediante la técnica aleatoria con 100.000 iteraciones sobre las cuales, los mejores resultados en cuanto al número de aciertos intra y extra muestral, estabilidad, significancia estadística DAT y beneficios al final del periodo extramuestral se encontraron en el modelo número 12, en que la predicción del signo de esta mejor combinación de funciones de activación y escalamiento, acertó 31,9% intra muestral y 32,8% extra muestral, con un acierto direccional de -3,99 estadísticamente significativo en un 5%. En dicho test, sobre un total de 116 observaciones (extra muestrales), la proyección de cambio de signos acertó 38 veces, lo que implica obtener un retorno equivalente al 565,7% sobre la inversión inicial en el evento de haber seguido de forma inversa a las recomendaciones, cifra que es ampliamente superior a la rentabilidad esperada de una estrategia “buy and hold”, la que alcanzo un retorno de 216,4%.

En el cuadro siguiente se muestra el resumen de los resultados obtenidos por esta metodología, destacando en color rojo el mejor modelo.

Porcentaje de Predicción de Signo (PPS)					
Modelo Red Neuronal	N° Iteración	Intra Muestral	Extra Muestral	Diferencial	Test DA
Red Ward RN Aleatorio Cat 1	100000	62.1%	64.7%	2.6%	3.21
Red Ward RN Aleatorio Cat 2	100000	34.5%	36.2%	1.7%	-2.85
Red Ward RN Aleatorio Cat 3	100000	37.1%	37.1%	0.0%	-2.59
Red Ward RN Aleatorio Cat 4	100000	37.1%	33.6%	-3.5%	-3.55
Red Ward RN Aleatorio Cat 5	100000	36.3%	33.6%	-2.7%	-3.46
Red Ward RN Aleatorio Cat 6	100000	35.3%	34.5%	-0.8%	-3.23
Red Ward RN Aleatorio Cat 7	100000	36.2%	36.2%	0.0%	-2.83
Red Ward RN Aleatorio Cat 8	100000	37.1%	37.1%	0.0%	-2.85
Red Ward RN Aleatorio Cat 9	100000	37.1%	34.5%	-2.6%	-4.08
Red Ward RN Aleatorio Cat 10	100000	36.2%	36.2%	0.0%	-2.83
Red Ward RN Aleatorio Cat 11	100000	33.6%	36.2%	2.6%	-2.81
Red Ward RN Aleatorio Cat 12	100000	31.9%	32.8%	0.9%	-3.99

Cuadro N° 5

Resultado del porcentaje de predicción de signos (PPS) de los modelos estudiados y el test de Pesaran y Timmenmann.



Retorno por Estrategia Periodo Extra Muestral				
Modelo Red Neuronal	N° Iteración	Estrategia		Diferencial
		Trading	Buy and Hold	
Red Ward RN Aleatorio Cat 1	100000	428.3%	216.4%	211.9%
Red Ward RN Aleatorio Cat 2	100000	486.3%	216.4%	270.0%
Red Ward RN Aleatorio Cat 3	100000	219.2%	216.4%	2.8%
Red Ward RN Aleatorio Cat 4	100000	355.3%	216.4%	139.0%
Red Ward RN Aleatorio Cat 5	100000	360.9%	216.4%	144.5%
Red Ward RN Aleatorio Cat 6	100000	381.2%	216.4%	164.9%
Red Ward RN Aleatorio Cat 7	100000	369.9%	216.4%	153.5%
Red Ward RN Aleatorio Cat 8	100000	343.4%	216.4%	127.0%
Red Ward RN Aleatorio Cat 9	100000	438.1%	216.4%	221.7%
Red Ward RN Aleatorio Cat 10	100000	343.7%	216.4%	127.3%
Red Ward RN Aleatorio Cat 11	100000	238.3%	216.4%	21.9%
Red Ward RN Aleatorio Cat 12	100000	565.7%	216.4%	349.3%

Cuadro N° 6

Resultado de los retornos del modelo v/s Estrategia Buy and Hold.

5.2. Modelo Multivariable Lineal

En esta metodología, básicamente replicamos las mismas variables de los modelos de redes neuronales, con la diferencia de que se aplicaron funciones lineales. Luego de iterar 50.000 y con un máximo de 100.000 veces los 12 modelos, los mejores resultados en cuanto al número de aciertos intra y extra muestral, estabilidad, significancia estadística DAT y beneficios al final del periodo extramuestral se encontraron en el modelo número 6 en que la predicción del signo de esta mejor combinación de funciones de activación y escalamiento, acertó 36,2% intra muestral y 33,6% extra muestral, con un acierto direccional de -3,43 estadísticamente significativo en un 5%. En dicho test, sobre un total de 116 observaciones (extra muestrales), la proyección de cambio de signos acertó 39 veces, lo que implica obtener un retorno equivalente al 412,1% sobre la inversión inicial en el evento de haber seguido de forma inversa a las recomendaciones, cifra que es ampliamente superior a la rentabilidad esperada de una estrategia "buy and hold", la que alcanzo un retorno de 216,4%.





En los cuadros 7 y 8 siguientes se muestra el resumen de los resultados obtenidos por esta metodología, destacando en color rojo el mejor modelo.

Porcentaje de Predicción de Signo PPS					
Modelo Multivariado Lineal	N° Iteración	Intra Muestral	Extra Muestral	Diferencial	Test DA
Multivariado Lineal Cat 1	100000	59.5%	59.5%	0.0%	2.25
Multivariado Lineal Cat 2	50000	37.1%	38.8%	1.7%	-2.27
Multivariado Lineal Cat 3	50000	39.7%	39.7%	0.0%	-1.98
Multivariado Lineal Cat 4	50000	37.9%	37.9%	0.0%	-2.72
Multivariado Lineal Cat 5	50000	37.1%	38.8%	1.7%	-2.21
Multivariado Lineal Cat 6	50000	36.2%	33.6%	-2.6%	-3.43
Multivariado Lineal Cat 7	50000	39.7%	38.8%	-0.9%	-2.19
Multivariado Lineal Cat 8	50000	62.1%	62.1%	0.0%	2.63
Multivariado Lineal Cat 9	50000	38.8%	37.9%	-0.9%	-2.45
Multivariado Lineal Cat 10	50000	37.9%	37.9%	0.0%	-2.50
Multivariado Lineal Cat 11	50000	38.8%	37.1%	-1.7%	-2.85
Multivariado Lineal Cat 12	50000	37.1%	38.8%	1.7%	-2.16

Cuadro N° 7

Resultado del porcentaje de predicción de signos (PPS) de los modelos estudiados y el Test de Pesaran y Timmenrmann.

Retorno por Estrategia Periodo Extra Muestral				
Modelo Multivariado Lineal	N° Iteración	Estrategia		Diferencial
		Trading	Buy and Hold	
Multivariado Lineal Cat 1	100000	243.9%	216.4%	27.5%
Multivariado Lineal Cat 2	50000	240.3%	216.4%	23.9%
Multivariado Lineal Cat 3	50000	170.9%	216.4%	-45.4%
Multivariado Lineal Cat 4	50000	250.4%	216.4%	34.0%
Multivariado Lineal Cat 5	50000	236.3%	216.4%	19.9%
Multivariado Lineal Cat 6	50000	412.1%	216.4%	195.7%
Multivariado Lineal Cat 7	50000	310.8%	216.4%	94.5%
Multivariado Lineal Cat 8	50000	320.0%	216.4%	103.6%
Multivariado Lineal Cat 9	50000	168.8%	216.4%	-47.5%
Multivariado Lineal Cat 10	50000	213.5%	216.4%	-2.9%
Multivariado Lineal Cat 11	50000	333.6%	216.4%	117.2%
Multivariado Lineal Cat 12	50000	293.1%	216.4%	76.7%

Cuadro N° 8

Resultado de los retornos del modelo v/s Estrategia Buy and Hold.



5.3. Convergencia de los Modelos

Con respecto de la metodología de iteraciones de la función aleatoria entre -1 a +1 de los ponderadores (Alfas), se puede establecer como una aproximación empírica que la mayor productividad de esta metodología ocurre antes de las 40.000 iteraciones, siendo marginal el incremento del porcentaje de aciertos entre las 40.000 y 100.000 iteraciones.

Los resultados mencionados se muestran en la gráfica siguiente con una función logarítmica equivalente.

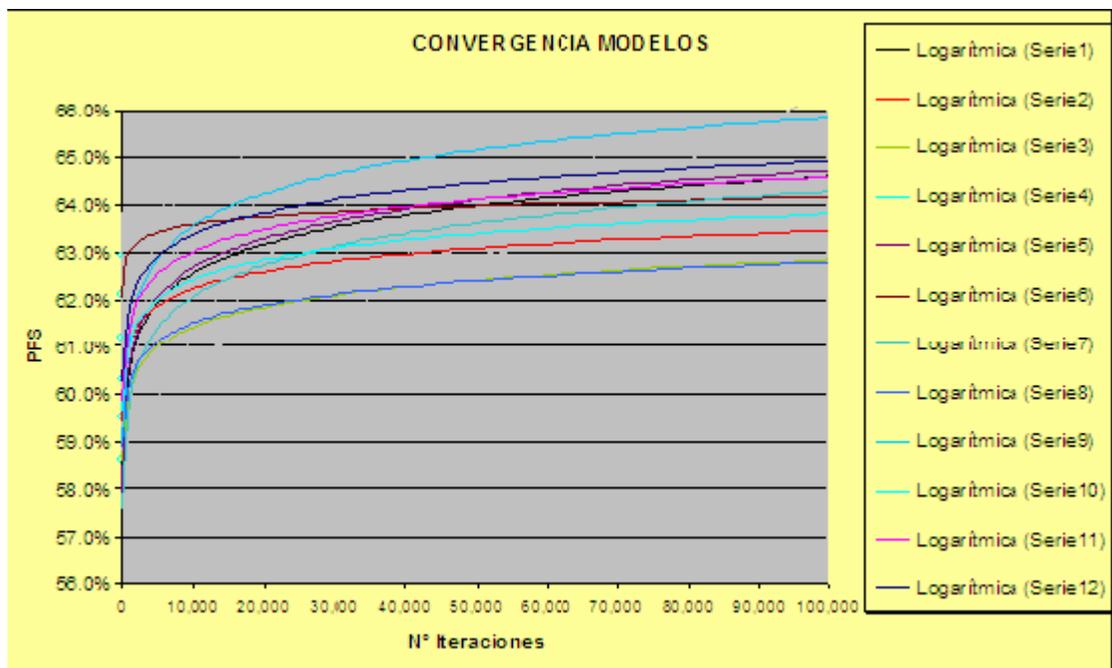


Grafico N° 3



6. CONCLUSIONES

Con los resultados obtenidos, después de testear los distintos modelos y metodologías, concluimos que la aplicación de Redes Neuronales Ward y sin necesidad de levantar supuestos sobre el comportamiento de los retornos, logran predecir el cambio de signo en el precio de la acción Caterpillar a tres semanas.

Las Redes Neuronales Ward superan el porcentaje de predicción de signos de los modelos Multivariados Lineales.

La prueba de acierto direccional (DAT) de Pesaran & Timmermann (1992), para el mejor modelo de Red Neuronal Ward, con un intervalo de confianza del 95% (5% de significancia con valor crítico de 1,96), arrojó un valor de -3.99, con lo que inferimos que el modelo es estadísticamente significativo y tiene capacidad predictiva.

La mejor rentabilidad mostrada para la metodología de Red Neuronal Ward de una inversión ficticia de US\$10.000, siguiendo las recomendaciones predictivas del modelo (compra o venta), se obtuvo un beneficio de 565,7%, superando ampliamente a los retornos que se hubiesen obtenido con la estrategia "Buy and Hold", en cuyo caso el retorno de la inversión alcanzó 216,4%.

La predicción a tres semanas tiene la ventaja intrínseca de disminuir las transacciones, con lo cual se disminuyen los costos de transacción.

Finalmente, podemos afirmar existe evidencia significativa de que los precios accionarios no siguen un camino aleatorio y muestran que los retornos accionarios son predecibles en algún grado, en nuestro caso, con un modelo inverso con acierto extra muestral del 32,8% (equivalente al 67,2%), eliminando la hipótesis de que el precio de la acción sigue un comportamiento de "caminata aleatoria".



7. BIBLIOGRAFÍA Y REFERENCIAS:

- Alvarez-Diaz, M. & A. Alvarez (2002) Forecasting exchange rates using genetic algorithms, Applied Economic Letters, forthcoming.
- Baille R. & P.McMahon (1989) The foreign exchange market: Theory and evidence. Cambridge University Press.
- Ball, R. & Kothari, S.P. (1989). Nonstationary expected returns: Implications for test of market efficiency and serial correlation in returns. Journal of Financial Economics 25, 51-74.
- Blume, L., Easley D. & O'Hara M. (1994). Market statistics and technical analysis: The role of volume. Journal of Finance 49, 153-82.
- Brock, W., Lakonishok J. & LeBaron B. (1992). Simple technical trading rules and the stochastic properties of stocks returns. The Journal of Finance, Vol. XLVII, N°5, 1731-1763 (December).
- Chan, K.C. (1988). On the contrarian investment strategy. Journal of Business 61, 147-163.
- Chopra, N., Lakonishok J. & Ritter J. R. (1992). Measuring abnormal returns: Do stocks overreact? Journal of Financial Economics 31, 235-268.
- Conrad J. & Kaul G. (1988). Time-variation in expected returns. Journal of Business 61, 409-425.
- Conrad J. & Kaul G. (1989). Mean reversion in short-horizon expected returns. Review of Financial Studies 2, 225-240.
- Conrad J. & Kaul G. (1993). Long-term market overreaction or biases in computed return? Journal of Finance 48, 39-63.
- DeBondt, W. F. M. & Thaler R. (1985). Does the stock market overreact? Journal of Finance 40, 793-805.
- DeBondt, W. F. M. & Thaler R. (1987). Further evidence on investor overreaction and stock market seasonability. Journal of Finance 42, 557-581.
- Diebold, F.X. & J.A Nason (1990) Nonparametric exchange rate prediction?, Journal of International Economics, 28, pp. 315-332



Diaz, D (2005) Prediciendo el Cambio de Signo del Precio del Oro: Modelos de Redes Neuronales Rolling y Recursivos. Tesis Facultad de Ciencias Economicas y Administrativas, Universidad de Chile.

Ferson, W. & Harvey, C. (1991). The variation of economic risk premiums. *Journal of Political Economy* 99, 385-415.

French, K. R. & Roll R. (1986). Stock return variances: The arrival of information and reaction of traders. *Journal of Financial Economics* 17, 5-26.

Friz, R (2005) Determinacion de Optimo de Rolling, en Modelos ARIMAX. Tesis Facultad de Ciencias Economicas y Administrativas, Universidad de Chile.

Hu M.Y.,G.Zhang,C.X.Jiang & B.E Patuwo (1999) A cross – validation analisis of neural network out-ofsample performance in exchange rate forecasting. *Decision Sciencies*, 30,1, 197-216.

Grundy, B. & McNichols M. (1989). Trade and the revelation of information through prices and direct disclosure. *Review of Financial Studies* 2, 495-526.

Goldberg D.E (1989) Genetic algorithms in search, optimization and machine learning. Reading, MA: Addison – Wesley.

Gonzáles I. & Jiménez J.(2003) Redes neuronales aplicadas a la prediccion del precio del tipo de cambio, Rolling vs. Recursivo. Tesis Facultad de Ciencias Economicas y Administrativas, Universidad de Chile.

Holland J.H (1975) Adaptation in natural and artificial systems, Ann Arbor. The University of Michigan Press.

Hsieh, D.A (1989) Testing for nolinear dependence in daily foreign exchange rates, *Journal of Business*, 62, pp. 329-368.

Jegadeesh, Narasimhan (1990). Evidence of predictable behavior of security returns. *The Journal of*



Finance, Vol. 45(3), 881-898 (July).

Jensen, Michael C. & Benington, George A (May 1970). Random Walks and Technical Theories: Some

Additional Evidence. Journal of Finance. Reprinted in Security Evaluation and Portfolio Analysis, E. Elton

and M. Gruber, Editors (Prentice-Hall, 1972), and Investment Management: Some Readings, J. Lorie and

R. Brealey, Editors (Praeger Publishers, 1972).

Kuan C.M. & T.Liu (1995) Forecasting exchange rates using feedforward and recurrent neural network.

Journal of Applied Econometrics, 10, 4, 347 – 364.

Lehmann, B. (1990) Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1-28.

Lisi F. & A. Medio (1997) Is a random walk the best exchange rate predictor?, International Journal of

Forecasting, 13, 255-267

Tesis para optar al Grado de Magíster en Finanzas

40

Lo, A. & MacKinley A. C. (1988). Stock market price do not follow random walk: Evidence from a simple

specification test. Review of Financial Studies 1, 41-66.

Malkiel, B. (1981). A random walk down Wall Street, 2 ed. (Norton, New York).

Pesaran M.H. & Timmermann (1992) A simple no parametric test of predictive performance. Journal of

Business & Economic Statistics, 10,4, 461- 465.

Parisi, F. & Vasquez, A. (2000). Simple technical trading rules of stock returns: evidence from 1987 to

1998 in Chile. Emerging Markets Review, Vol. 1, N°2, 152-164.

Parisi, A. (2002) "Evaluacion de modelos de redes neuronales de prediccion del signo de la variacion del IPSA", Estudios de Administracion, vol. 9, No 1, Universidad de Chile, 2002, p. 76.

Parisi, Parisi (2003) Modelos de redes neuronales Rolling versus Recursivo: Precio del Oro, Facultad de

Ciencias Economicas y Administrativas, Universidad de Chile.

Parisi A., Parisi F. & E.Cornejo (2003) "Algoritmos Geneticos: Prediccion de Indices Bursatiles de



America del Norte. Universidad de Chile

Poterba, J. M. & Summers L. H. (1988). Mean reversion in stock prices: Evidence and implications.

Journal of Financial Economics 22, 27-59.

Shefrin, H. M. & Statman M. (1985). The disposition to ride winners too long and sell losers too soon:

Theory and evidence. Journal of Finance 41, 774-790.

Urrutia, J. (1994). Time series properties of four Latin American equity markets: Argentina, Brazil, Chile

and Mexico. Estudios de Administracion 2, 1-8.

Vega, M. & Alvarez, J. (2001) Tipos de cambio flexibles y volatilidad: las regularidades empiricas de las

observaciones diarias. Universidad de Navarra.

Vera, L. (2005) Prediccion del cambio de signos en los tipos de cambio Peso chileno y Peso Mexicano.

Tesis Facultad de Ciencias Economicas y Administrativas, Universidad de Chile.

Inouye, G. (2005) Aplicacion de redes neuronales y Algoritmos Geneticos para predecir la apreciacion o

depreciacion de los tipos de cambio, evidencias para el Real Brasileiro y Nuevo Sol Peruano.

Tesis

Facultad de Ciencias Economicas y Administrativas, Universidad de Chile.

Zarowin, P. (1990). Size, seasonality, and stock market overreaction. Journal of Financial and Quantitative Analysis 25, 113-125.



ANEXOS:

ANEXO 1.- Resultados y Ponderadores Mejor Modelo Red Neuronal

Ponderadores Slab-2 Mejor Modelo				
Sesgo	Var1	Var2	Var3	Var4
-0.139799	0.626273	-0.953790	-0.746766	0.437984
-0.590625	0.583821	0.126899	0.693818	0.411057
0.014920	-0.667421	-0.314178	0.880719	-0.034865
0.284437	-0.350458	0.574804	-0.422893	-0.814391
0.234209	-0.214120	-0.452409	-0.690136	0.598291
0.860521	0.292218	0.293878	0.597441	-0.107684

Ponderadores Slab - 3 Mejor Modelo				
Sesgo	Var1	Var2	Var3	Var4
-0.303745	-0.744890	0.853110	0.820528	0.826522
0.831837	-0.048657	0.566249	-0.446340	0.921740
-0.354626	-0.854174	-0.422801	0.093318	0.900644
0.746613	-0.698922	0.314486	0.986275	0.626446
0.486140	0.689990	0.811646	-0.405213	0.821308
0.072779	-0.002965	-0.701141	0.512128	0.032955

Ponderadores Slab - 4 Salida Mejor Modelo						
Sesgo	N1	N2	N3	N4	N5	N6
-0.402318	-0.539033	-0.104785	0.952370	-0.121602	0.171404	0.589947
-0.815363	-0.652731	-0.108673	-0.359344	0.910786	-0.752646	0.799672

Date	CAT Real	MODELO Datos de Entrada (X)				Slab 1 - Entrada Escalamiento TanH			
		BHP (-2)	PH (-2)	GSPC (-1)	CH (-1)	Ent.1	Ent.2	Ent.3	Ent.4
08-Ene-93	0%	0%	0%	0%	0%	-0.142047	-0.159592	-0.157965	-0.118602
29-Ene-93	-1.31%	0%	0%	0%	0%	-0.142047	-0.159592	-0.157965	-0.118602
19-Feb-93	-0.76%	0%	0%	2.27%	13.79%	-0.142047	-0.159592	0.476306	0.955508
12-Mar-93	3.82%	2.91%	0.31%	-1.04%	-4.55%	0.252882	-0.117664	-0.438012	-0.653728
02-Abr-93	0.92%	9.58%	-1.53%	3.59%	1.79%	0.827195	-0.357570	0.723369	0.140222
23-Abr-93	14.96%	6.88%	1.24%	-1.88%	-3.12%	0.667645	0.012184	-0.616800	-0.518163
14-May-93	3.81%	-0.94%	-5.83%	-0.99%	-4.02%	-0.266017	-0.750170	-0.425505	-0.608029
04-Jun-93	7.65%	2.44%	-3.80%	0.58%	-9.85%	0.190921	-0.598432	0.013647	-0.914694
25-Jun-93	-2.84%	-4.23%	7.91%	2.39%	-0.47%	-0.621027	0.735404	0.503758	-0.184817
16-Jul-93	7.60%	-1.38%	9.74%	-0.55%	12.38%	-0.321625	0.832129	-0.311848	0.933637
06-Ago-93	1.09%	-1.54%	-0.76%	-0.41%	-3.12%	-0.341185	-0.261075	-0.275471	-0.518120

SLAB2 - Capa Oculta Activacion Gaussiana					
N1	N2	N3	N4	N5	N6
0.999889	0.483757	0.999381	0.848119	0.868891	0.620765
0.999889	0.483757	0.999381	0.848119	0.868891	0.620765
0.999811	0.999134	0.742180	0.580819	0.714519	0.402643
0.970992	0.345760	0.794319	0.489006	0.979347	0.605294
0.943789	0.847716	0.957909	0.671472	0.962134	0.135329
0.778516	0.493617	0.398135	0.578090	0.960248	0.572779
0.808877	0.146346	0.994450	0.679524	0.730458	0.869734
0.980663	0.427675	0.985840	0.687019	0.991461	0.487936
0.058022	0.709087	0.656708	0.475441	0.835630	0.227926
0.784152	0.774593	0.891401	0.940816	0.612270	0.591844
0.984329	0.221873	0.990028	0.533722	0.910891	0.718217

SLAB3 - Capa Oculta Activacion ComGausiana					
N1	N2	N3	N4	N5	N6
0.270603	0.395578	0.079267	0.273788	0.049449	0.010008
0.270603	0.395578	0.079267	0.273788	0.049449	0.010008
0.511560	0.865543	0.420981	0.969028	0.514757	0.191096
0.892117	0.112729	0.733859	0.090874	0.041430	0.080307
0.233378	0.144797	0.401286	0.520477	0.292908	0.384039
0.948993	0.305626	0.879366	0.343845	0.456680	0.070658
0.922014	0.002445	0.146192	0.107046	0.330344	0.122562
0.944670	0.125061	0.693355	0.017946	0.323042	0.197181
0.665742	0.541674	0.062734	0.959839	0.085304	0.035128
0.740534	0.995370	0.134489	0.897850	0.965233	0.334933
0.576235	0.112819	0.179768	0.089780	0.072880	0.009692

SLAB 4 - Capa Salida Activacion TanH					
S1	Prediccion	Real	Error	Error^2	pps
-0.439412	-0.027792	0.000000	0.027792	0.000772	0
-0.439412	-0.027792	-0.013109	0.014683	0.000216	1
-0.662010	-0.088281	-0.007590	0.080691	0.006511	1
-0.875316	-0.146245	0.038241	0.184486	0.034035	0
-0.455354	-0.032124	0.009208	0.041332	0.001708	0
-0.948875	-0.166234	0.149635	0.315869	0.099773	0
-0.745050	-0.110846	0.038095	0.148942	0.022184	0
-0.890934	-0.150489	0.076453	0.226942	0.051503	0
-0.154564	0.049614	-0.028409	-0.078023	0.006088	0
-0.671598	-0.090887	0.076023	0.166910	0.027859	0
-0.624300	-0.078033	0.010870	0.088903	0.007904	0

TIEMPO DE EJECUCION	
hora inicio	15-08-2006 13:20
hora término	15-08-2006 15:22
Total	2:01:55

OBJETIVOS	
PPS Objetivo	55%
Iteraciones	100000
Avance Itera	100000

RESULTADOS		
INVERSO	ENCONTRADO	MEJOR PPS
68.1%	31.9%	Intramuestral
67.2%	32.8%	Extramuestral



ANEXO 2. Fundamentos de las Redes Neuronales

El Modelo Biológico

Se estima que el cerebro humano contiene más de cien mil millones de neuronas estudios sobre la anatomía del cerebro humano concluyen que hay más de 1000 sinápsis a la entrada y a la salida de cada neurona. Es importante notar que aunque el tiempo de conmutación de la neurona (*unos pocos milisegundos*) es casi un millón de veces menor que en los actuales elementos de las computadoras, ellas tienen una conectividad miles de veces superior que las actuales supercomputadoras.

Las neuronas y las conexiones entre ellas (sinápsis) constituyen la clave para el procesamiento de la información.

Algunos elementos ha destacar de su estructura histológica son:

Las dendritas, que son la vía de entrada de las señales que se combinan en el cuerpo de la neurona. De alguna manera la neurona elabora una señal de salida a partir de ellas.

El axón, que es el camino de salida de la señal generada por la neurona.

Las sinapsis, que son las unidades funcionales y estructurales elementales que median entre las interacciones de las neuronas. En las terminaciones de las sinapsis se encuentran unas vesículas que contienen unas sustancias químicas llamadas neurotransmisores, que ayudan a la propagación de las señales electroquímicas de una neurona a otra.

Lo que básicamente ocurre en una neurona biológica es lo siguiente: la neurona es estimulada o excitada a través de sus entradas (inputs) y cuando se alcanza un cierto umbral, la neurona se dispara o activa, pasando una señal hacia el *axón*.

Posteriores investigaciones condujeron al descubrimiento de que estos procesos son el resultado de eventos electroquímicos. Como ya se sabe, el pensamiento tiene lugar en el cerebro, que consta de billones de neuronas interconectadas. Así, el secreto de la "inteligencia" -sin importar como se defina- se sitúa dentro de estas neuronas interconectadas y de su interacción.



La forma que dos neuronas interactúan no está totalmente conocida, dependiendo además de cada neurona. En general, una neurona envía su salida a otras por su axón. El axón lleva la información por medio de diferencias de potencial, u ondas de corriente, que depende del potencial de la neurona.

Este proceso es a menudo modelado como una regla de propagación representada por la función de red $u(.)$. La neurona recoge las señales por su sinápsis sumando todas las influencias excitadoras e inhibitoras. Si las influencias excitadoras positivas dominan, entonces la neurona da una señal positiva y manda este mensaje a otras neuronas por sus sinápsis de salida. En este sentido la neurona puede ser modelada como una simple función escalón $f(.)$. Como se muestra en la próxima figura, la neurona se activa si la fuerza combinada de la señal de entrada es superior a un cierto nivel, en el caso general el valor de activación de la neurona viene dado por una función de activación $f(.)$.

Otras definiciones de Redes Neuronales

Una red neuronal es un procesador masivamente paralelo distribuido que es propenso por naturaleza a almacenar conocimiento experimental y hacerlo disponible para su uso. Este mecanismo se parece al cerebro en dos aspectos [SH1]:

El conocimiento es adquirido por la red a través de un proceso que se denomina aprendizaje. El conocimiento se almacena mediante la modificación de la fuerza o peso sináptico de las distintas uniones entre neuronas.

Una red neuronal es un **modelo computacional** con un conjunto de propiedades específicas, como son la habilidad de adaptarse o aprender, generalizar u organizar la información, todo ello basado en un procesamiento eminentemente paralelo. [BJAK] [PPVS]

Elementos de una Red Neuronal Artificial

Las redes neuronales son modelos que intentan reproducir el comportamiento del cerebro. Los mismos constan de dispositivos elementales de proceso: *las neuronas*. A partir de ellas, se pueden generar representaciones específicas, de tal forma que un estado conjunto de ellas



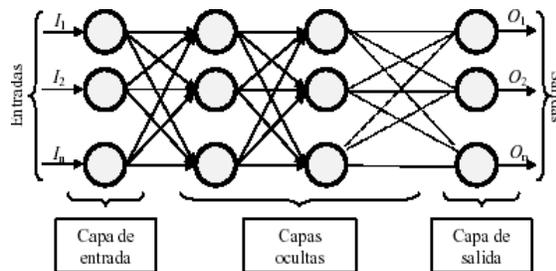
puede significar una letra, un número u otro objeto. Generalmente se pueden encontrar tres tipos de neuronas:

Aquellas que reciben estímulos externos relacionados con el aparato sensorial, que tomarán la información de entrada.

Dicha información se transmite a ciertos elementos internos que se ocupan de su procesamiento. Es en las *sinapsis* y *neuronas* correspondientes a este segundo nivel donde se genera cualquier tipo de representación interna de información. Como no tienen relación directa con la información de entrada ni con la salida, estos elementos se denominan *unidades ocultas*.

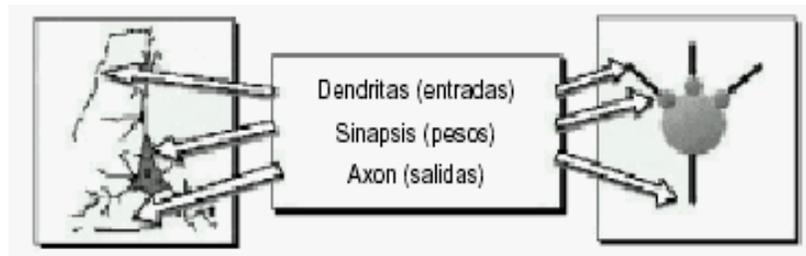
Una vez finalizado el período de procesado, la información llega a las unidades de salida, cuya misión es dar la respuesta al sistema.

A continuación se puede ver en la siguiente figura, un esquema de una red neuronal:



La misma está constituida por neuronas interconectadas y arregladas en tres capas (esto último puede variar). Los datos ingresan por medio de la "capa de entrada", pasan a través de la "capa oculta" y salen por la "capa de salida". Cabe mencionar que la capa oculta puede estar constituida por varias capas.

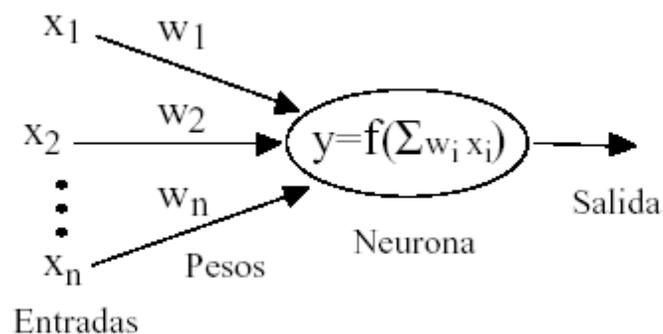
En la siguiente figura se compara una neurona biológica con una neurona artificial. En la misma se pueden observar las similitudes entre ambas (tienen entradas, utilizan pesos y generan salidas).



La neurona artificial pretende mimetizar las características más importantes de las neuronas biológicas. Cada neurona i -ésima está caracterizada en cualquier instante por un valor numérico denominado valor o estado de activación $a_i(t)$; asociado a cada unidad, existe una función de salida, f_i , que transforma el estado actual de activación en una señal de salida. Dicha señal es enviada a través de los canales de comunicación unidireccionales a otras unidades de la red; estos canales la señal se modifica de acuerdo con la sinápsis (el peso, w_{ji}) asociada a cada uno de ellos según determinada regla. Las señales moduladas que han llegado a la unidad j -ésima se combinan entre ellas, generando así la entrada total Net_j .

$$Net_j = \sum_i y_i w_{ji}$$

Una función de activación, F , determina el nuevo estado de activación $a_j(t+1)$ de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación $a_j(t)$.



La dinámica que rige la actualización de los estados de las unidades puede ser de dos tipos: asíncrono y modo síncrono. En el primer caso, las neuronas evalúan su estado continuamente según les va llegando información, y lo hacen de forma independiente, En el segundo caso, la



información llega de forma continua, pero los cambios se realizan simultáneamente, como si existiera un reloj interno que decidiera cuando cambiar su estado. Los sistemas biológicos quedan probablemente entre ambas posibilidades.

Unidades de Proceso: La Neurona Artificial

Si se tienen N unidades (neuronas), podemos ordenarlas arbitrariamente y designar la j-ésima unidad como U_j . Su trabajo es simple y único, y consiste en recibir las entradas de las células vecinas y calcular un valor de salida, el cual es enviado a todas las células restantes.

En cualquier sistema que se esté modelando, es útil caracterizar tres tipos de unidades: entradas, salidas y ocultas. Las unidades de entrada reciben señales del entorno, éstas pueden ser provenientes de sensores o de otros sectores del sistema. Las unidades de salida envían la señal fuera del sistema; éstas pueden controlar directamente potencias u otros sistemas. Las unidades ocultas son aquellas cuyas entradas y salidas se encuentran dentro del sistema; es decir no tienen contacto con el exterior.

Se conoce como nivel o capa a un conjunto de neuronas cuyas entradas provienen de la misma fuente, y cuyas salidas se dirigen a un mismo destino.

Estado de Activación

Junto al conjunto de unidades, la representación necesita los estados del sistema en un tiempo t . Esto se especifica en un vector de N números reales $A(t)$, que representa el *estado de activación* del conjunto de unidades de procesamiento. Cada elemento del vector representa la activación de una unidad en el tiempo t . La activación de una unidad U_i en el tiempo t se designa por $a_i(t)$; es decir:

$$A(t) = (a_1(t), \dots, a_i(t), \dots, a_N(t))$$

El procesamiento que realiza la red se ve como la evolución de un patrón de activación en el conjunto de unidades que lo componen a través del tiempo.

Todas las neuronas que componen la red se hallan en cierto estado. Podemos decir que hay dos posibles estados, *reposo* y *excitado*, a los que denominaremos *estados de activación* y a cada uno de los cuales se le asigna un valor. Los valores de activación pueden ser continuos o discretos. Además pueden ser limitados o ilimitados. Si son discretos, suelen tomar un conjunto



pequeño de valores o bien valores binarios. En notación binaria, un estado activo se indicaría por un 1, y se caracteriza por la emisión de un impulso por parte de la neurona (potencial de acción), mientras que un estado pasivo se indicaría por un 0. En otros modelos se considera un conjunto continuo de estados de activación, en cuyo caso se asigna un valor entre $[0,1]$ o en el intervalo $[-1,1]$, generalmente siguiendo una función sigmoïdal.

Los criterios o reglas que siguen las neuronas para alcanzar tales estados dependen de dos factores:

Dado que las propiedades macroscópicas de las redes neuronales no son producto de actuación de elementos individuales, es necesario tener idea del mecanismo de interacción entre las neuronas. El estado de activación estará fuertemente influenciado por tales interacciones ya que el efecto que producirá una neurona sobre otra será proporcional a la fuerza, peso de la conexión entre ambas.

La señal que envía cada una de las neuronas a sus vecinas dependerá de su propio estado de activación.

Función de Salida o Transferencia

Asociada a cada unidad U_i (neurona) hay una función de salida $f_i(a_i(t))$, que transforma el estado actual de activación en una señal de salida:

En algunos modelos, esta salida es igual al nivel de activación de la unidad, en cuyo caso la función f_i es la función identidad, $f_i(a_i(t)) = a_i(t)$. A menudo, f_i es de tipo sigmoïdal, y suele ser la misma para todas las unidades.

Existen cuatro funciones de transferencia típicas que determinan distintos tipos de neuronas:

- ✓ *Función Escalón*
- ✓ *Función Lineal y Mixta*
- ✓ *Sigmoïdal*
- ✓ *Función Gaussiana*

La función escalón únicamente se utiliza cuando las salidas de la red son binarias. La salida de una neurona se activa sólo cuando el estado de activación es mayor o igual a cierto valor



umbral. La función lineal o identidad equivale a no aplicar función de salida. Las funciones mixta y sigmoideal son las más apropiadas cuando queremos como salida información analógica.

Neurona de Función Escalón

La función escalón se asocia a neuronas binarias en las cuales cuando la suma de las entradas es mayor o igual que el umbral de la neurona, la activación es 1, si es menor, la activación es 0 (ó -1). Las redes formadas por este tipo de neuronas son fáciles de implementar en hardware, pero sus capacidades están limitadas.

Neurona de Función Lineal o Mixta

La función lineal o mixta corresponde a la función $F(x) = x$. En las neuronas con función mixta si la suma de las señales de entrada es menor que un límite inferior, la activación se define como 0 (ó -1). Si dicha suma es mayor o igual que el límite superior, entonces la activación es 1. Si la suma de entrada está comprendida entre ambos límites, la activación se define como una función lineal de suma de las señales de entrada.

Neurona de Función Continua (Sigmoideal)

Cualquier función definida simplemente en un intervalo de posibles valores de entrada, con un incremento monótonico y que tengan ambos límites superiores e inferiores (por ejemplo las funciones sigmoideal y arco tangente), podrá realizar la función de activación o transferencia de forma satisfactoria.

Con la función sigmoideal, para la mayoría de los valores del estímulo de entrada, el valor dado por la función es cercano a uno de los valores asintóticos. Esto hace posible que en la mayoría de los casos, el valor de salida esté comprendido en la zona alta o baja del sigmoide. De hecho cuando la pendiente es elevada, esta función tiende a la función escalón. La importancia de ésta función es que su derivada es siempre positiva y cercana a cero para los valores grandes positivos o negativos; además toma su valor máximo cuando x es cero. Esto hace que se puedan utilizar las reglas de aprendizaje definidas para la función escalón, con la ventaja respecto a esta función, que la derivada está definida para todo el intervalo. La función escalón no podía definir la derivada en el punto de transición y esto no ayuda a los métodos de aprendizaje en los cuales se usan derivadas.



Función de Transferencia Gaussiana

Los centros y anchura de estas funciones pueden ser adaptados, lo cual las hace más adaptativas que las funciones sigmoideas.

Conexiones entre Neuronas

Las conexiones que unen las neuronas que forman una RNA tienen asociado un peso que es el que hace que la red adquiera conocimiento. Consideremos y_i como el valor de salida de una neurona i en un instante dado. Una neurona recibe un conjunto de señales que le dan información del estado de activación de todas las neuronas con las que se encuentra conectada. Cada conexión (sinápsis) entre la neurona i y la j está ponderada por un peso w_{ji} . Normalmente, como simplificación, se considera que el efecto de cada señal es aditivo, de tal forma que la entrada neta que recibe una neurona net_j , es la suma de cada señal individual por el valor de la sinapsis que conecta ambas neuronas:

$$net_j = \sum_i^N w_{ji} * y_i$$

Esta regla muestra el procedimiento a seguir para combinar los valores de entrada a una unidad con los pesos de las conexiones que llegan a esa unidad y es conocida como *regla de propagación*.

Función o Regla de Activación

Así como es necesario una regla que combine las entradas de una neurona con los pesos de las conexiones, también se requiere una regla que combine las entradas con el estado actual de la neurona para producir un nuevo estado de activación. Esta función F produce un nuevo estado de activación en una neurona a partir del estado (a_i) que existía y la combinación de las entradas con los pesos de las conexiones (net_i) .



Dado el estado de activación $a_i(t)$ de la unidad U_i y la entrada total que llega, Net_i , el estado de activación siguiente, $a_i(t+1)$, se obtiene aplicando una función F , llamada *función de activación*.

$$a_i(t+1) = F(a_i(t), Net_i)$$

En la mayoría de los casos la función F es la *función identidad*, por lo que el estado de activación de la neurona en $t+1$ coincidirá con el Net de la misma t . En este caso, el parámetro que se le pasa a la función de salida f , de la neurona será directamente el Net . Es estado de activación anterior no se tiene en cuenta. Según esto, la salida de una neurona y_i quedará según la expresión:

$$y_i(t+1) = f(Net_i) = f\left(\sum_{j=1}^N w_{ji} y_j(t)\right)$$

Por tanto, y en lo sucesivo, consideraremos únicamente la función f , que denominaremos de transferencia o de activación. Además, la misma no está centrada en el origen del eje que representa el valor de entrada neta sino que existe cierto desplazamiento debido a las características internas de la neurona y que no es igual en todas ellas. Este valor se denota como θ_i y representa el umbral de activación de la neurona i .

$$y_i(t+1) = f(Net_i - \theta_i) = f\left(\sum_{j=1}^N w_{ji} y_j(t) - \theta_i\right)$$

La salida se obtiene en una neurona para las diferentes forma de la función f serán:

Función de Activación Escalón

Si el conjunto de los estados de activación es $E = \{0, 1\}$, tenemos que:

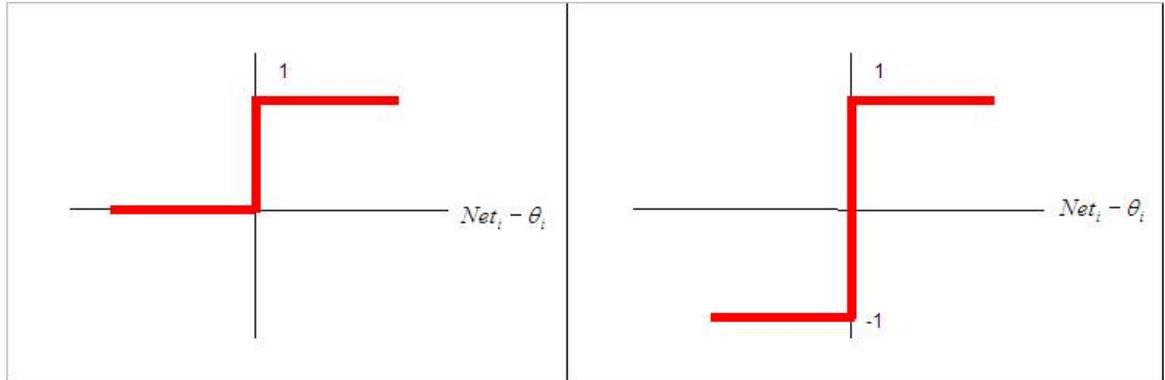
$$y_i(t+1) = \begin{cases} 1 & \text{si } [Net_i > \theta_i] \\ y(t) & \text{si } [Net_i = \theta_i] \\ 0 & \text{si } [Net_i < \theta_i] \end{cases}$$





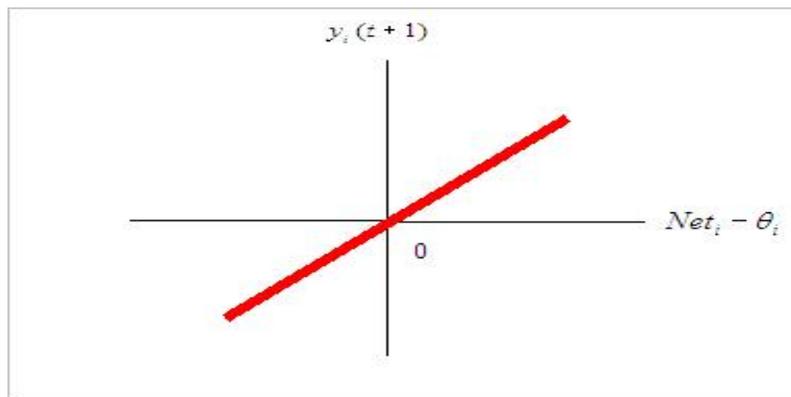
Si el conjunto es $E = \{-1, 1\}$, tendremos que:

$$y_i(t+1) = \begin{cases} 1 & \text{si } [Net_i > \theta_i] \\ y(t) & \text{si } [Net_i = \theta_i] \\ -1 & \text{si } [Net_i < \theta_i] \end{cases}$$



Función de Activación Lineal o Identidad

El conjunto de estados E puede contener cualquier número real; el estado de activación coincide con la entrada total que ha llegado a la unidad.

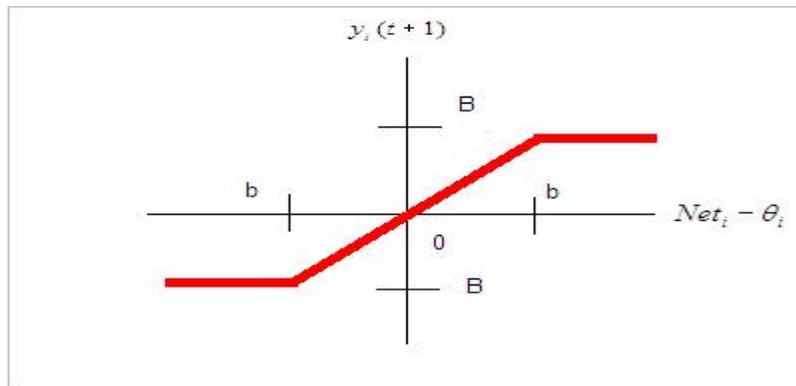




Función de activación lineal-mixta

$$y_i(t+1) = \begin{cases} b & \text{si } Net_i - \theta_i \leq -b \\ Net_i - \theta_i & \text{si } -b < Net_i - \theta_i < b \\ B & \text{si } Net_i - \theta_i \geq b \end{cases}$$

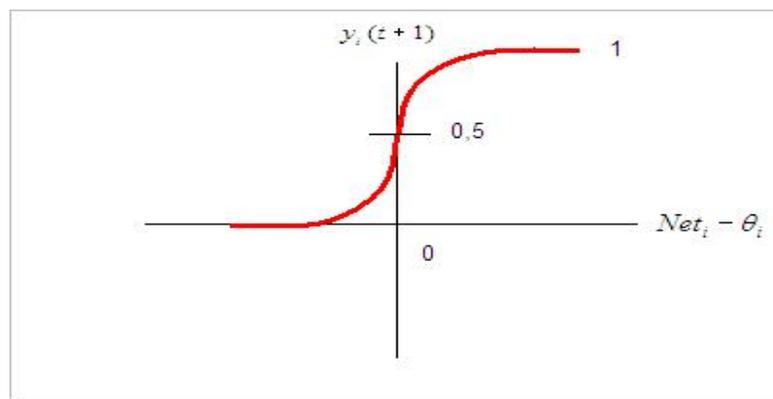
Con esta función, el estado de activación de la unidad está obligado a permanecer dentro de un intervalo de valores reales prefijados.



Función de activación sigmoideal

Es una función continua, por tanto el espacio de los estados de activación es un intervalo del eje real.

$$y_i(t+1) = \frac{1}{1 + e^{-x(Net_i - \theta_i)}}$$





Para simplificar la expresión de la salida de una neurona i , es habitual considerar la existencia de una neurona ficticia, con valor de salida unidad, asociada a la entrada de cada neurona i mediante una conexión con peso de valor $-\theta_i$. De esta forma la expresión de salida quedará:

$$y_i(t+1) = f\left(\sum_{j=1}^N w_{ji} y_j(t) - \theta_i * 1\right) = f\left(\sum_{j=1}^N w_{ji} y_j(t)\right) = f(Net_i)$$

Regla de aprendizaje

Existen muchas definiciones del concepto aprendizaje, una de ellas podría ser: La modificación del comportamiento inducido por la interacción con el entorno y como resultado de experiencias conducente al establecimiento de nuevos modelos de respuesta a estímulos externos.

Biológicamente, se suele aceptar que la información memorizada en el cerebro está mas relacionada con los valores sinápticos de las conexiones entre las neuronas que con ellas mismas. En el caso de las RNA, se puede considerar que el conocimiento se encuentra representado en los pesos de las conexiones entre neuronas. Todo proceso de aprendizaje implica cierto número de cambios en estas conexiones. Puede decirse que se aprende modificando los valores de los pesos de la red.

Estructura de una Red Neuronal Artificial

Niveles o capas de neuronas

La distribución de neuronas dentro de la red se realiza formando niveles o capas de un número determinado cada una. Se pueden distinguir tres tipos de capas:

De Entrada: es la capa que recibe directamente la información proveniente de las fuentes externas de la red.

Ocultas: son internas a la red, no tiene contacto directo con el exterior. El número de niveles ocultos puede ser de cero a un número elevado. Las neuronas de las capas ocultas pueden





estar interconectadas de distintas maneras, lo que determina junto a su número, las distintas topologías.

De Salida: transfieren información de la red hacia el exterior.

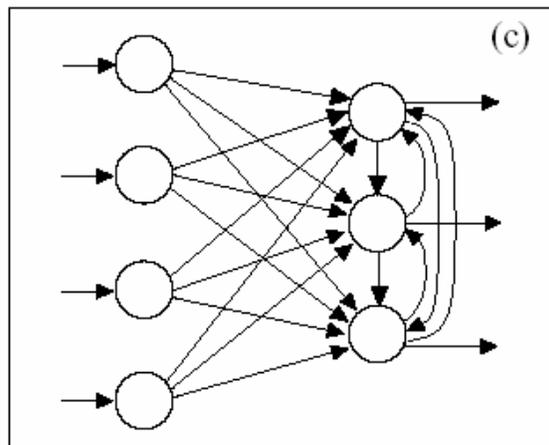
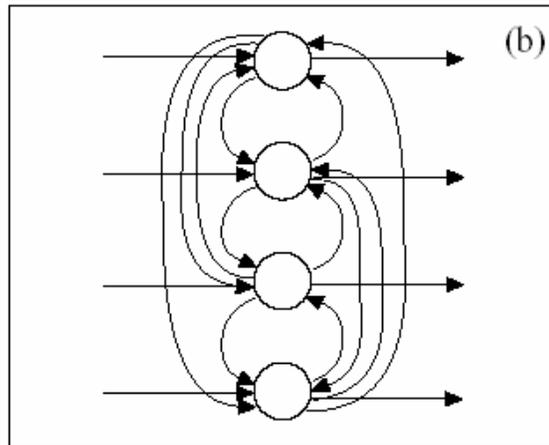
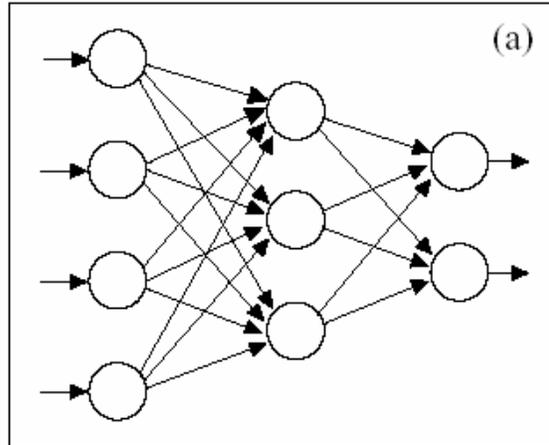
Se dice que una red es *totalmente conectada* si todas las salidas desde un nivel llegan a todos y cada uno de los nodos del mismo nivel siguiente.

Formas de Conexión entre neuronas

La conectividad entre los nodos de una red neuronal está relacionada con la forma en que las salidas de las neuronas están canalizadas para convertirse en entradas de otras neuronas. La señal de salida de un nodo puede ser una entrada de otro elemento de proceso, o incluso de sí mismo (conexión auto-recurrente).

Cuando ninguna salida de las neuronas es entrada de neuronas del mismo nivel o de niveles precedentes, la red se describe como propagación hacia delante. Cuando las salidas pueden estar conectadas como entradas de neuronas de niveles previos o del mismo nivel, incluyéndose ellas mismas, la red es de propagación hacia atrás. Las redes de propagación hacia atrás que tiene lazos cerrados son sistemas recurrentes.

En la siguiente figura se muestran ejemplos de conexiones.



- a) Conexiones hacia delante.
- b) Conexiones laterales.
- c) Conexiones hacia atrás (o recurrentes).



VENTAJAS DE LAS REDES NEURONALES ARTIFICIALES.

Adaptabilidad: Esta se basa en la capacidad que poseen las redes neuronales artificiales de aprender y realizar ciertas tareas basadas en un determinado entrenamiento o en una experiencia inicial. En este sentido, una red neuronal ya entrenada y operando puede realizar un aprendizaje continuo presentándole constantemente nuevos ejemplos. De esta forma, las redes neuronales pueden mejorar su capacidad de segregar entre entradas similares, adaptándose a los cambios del ambiente.

Autoorganización: A diferencia de las tecnologías actuales, en donde es necesario que toda la información sea especificada explícitamente por el programador, las redes neuronales, durante el proceso de aprendizaje o entrenamiento, organizan y crean su propia representación de la información que recibe. Así, las redes neuronales son adecuadas para problemas en los cuales es difícil o imposible definir un modelo explícito para obtener una determinada solución. Esta es una ventaja clara, ya que existe un gran número de problemas de este tipo.

Tolerancia a Fallas: En el caso de que una neurona o una porción completa de una red neuronal sean destruidas, esta continuara funcionando adecuadamente. Cuando el daño llega a ser tan extensivo que el comportamiento de la red comienza a ser afectado, el efecto será una degradación gradual de su rendimiento, y no una falla completa inmediata.

La razón de lo anterior radica en que las redes neuronales son capaces de almacenar la información de una manera distribuida. El conocimiento o información no es codificada ni manejada en una sola neurona, sino que a través de diversas neuronas y capas.

Procesamiento Rápido: Gracias a la capacidad de paralelizar el procesamiento de datos mediante la división de tareas computacionales en pequeñas subtareas (las cuales son repartidas sobre varios procesadores que operan simultáneamente), las redes neuronales aventajan a otras técnicas de procesamiento de datos, logrando uno de sus intereses más críticos: "La velocidad de procesamiento".

Compatibilidad con las tecnologías existentes: Como ocurre con la mayoría de las nuevas tecnologías, si los elementos necesarios para su uso fueran incompatibles y difíciles de integrar con otras tecnologías existentes y consolidadas, lo más probable, es que tendría muy poca



oportunidad de éxito y aceptación. Con esto en mente, las redes neuronales son sumamente prácticas, ya que pueden ser fácilmente integrables a casi cualquier tipo de sistema.

DESVENTAJAS DE LAS REDES NEURONALES ARTIFICIALES.

Las redes neuronales no permiten conocer a priori la topología con la que se va a solucionar del modo más eficiente un determinado problema. Si bien, existen métodos incipientes que tratan de solucionar de una manera formal el inconveniente anterior, aun son muy complejos, poco difundidos y no exactos, por lo que los métodos heurísticos cobran gran importancia.

Una vez entrenada una red neuronal, se hace difícil interpretar su funcionamiento, los datos entran y se obtienen predicciones, pero no se revela las relaciones entre las variables (Efecto “Black Box”). Aun mas, no es fácil asegurar con que grado de acierto responderá ante casos nunca vistos (generalización).

Ante un determinado problema de error de predicción, o sobre ajuste puede resultar dificultoso saber el por que del no funcionamiento adecuado de una red neuronal, ya que existen una multitud de factores que pueden influir en su rendimiento (Tasa de aprendizaje, momentum, numero de capas ocultas, numero de neuronas, escalamiento de datos, funciones de activación, entre otros).

Los Modelos Neuronales necesitan una herramienta de procesamiento poderosa, lo cual se manifiesta principalmente en el proceso de aprendizaje, pero este inconveniente se contrasta con la facilidad de implementación en dispositivos de hardware específicos.



ANEXO 3.-Test de acierto direccional (DA) de Pesaran & Timmermann (1992)

El test de acierto direccional³ se usa con el objetivo de medir la significancia estadística de la capacidad predictiva de los modelos analizados. El test de acierto direccional prueba la hipótesis nula de que las variaciones observadas están independientemente distribuidas de las variaciones proyectadas. Por ello, si se rechaza la hipótesis nula, se dice que existe evidencia estadística de que el modelo tiene capacidad para predecir la evolución futura de la variable observada.

Este test compara el signo de la proyección $\hat{\Delta P}_{i,t+1}^j$, con el del valor observado $\Delta P_{i,t+1}$ para cada i -ésima observación del conjunto extramuestral ($i=1,2,\dots,n$); donde el signo indica la dirección en que se moverá el mercado accionario: al alza, si es positivo, o a la baja, si es negativo. Si los signos coinciden, aumentan la efectividad de la predicción, y en caso de no existir coincidencia, aumenta el error de predicción del modelo (igual que la metodología usada para calcular PPS).

Luego, se calcula un ratio de éxito (SR) que se define como:

$$SR = n^{-1} \sum_{i=1}^m I_i [\Delta P_{i,t+1}, \hat{\Delta P}_{i,t+1}^j > 0]$$

donde $I_i[\cdot]$ es una función indicador que toma el valor de 1 cuando su argumento es cierto y 0 en otro caso. Además:

$$P = n^{-1} \sum_{i=1}^m I_i [\Delta P_{i,t+1} > 0]$$

$$\hat{P} = n^{-1} \sum_{i=1}^m I_i [\hat{\Delta P}_{i,t+1}^j > 0]$$

El ratio de éxito en el caso de independencia de $\hat{\Delta P}_{i,t+1}^j$ y $\Delta P_{i,t+1}$, SRI, está dada por:

³ Directional Accuracy Test.



$$SRI = P \cdot \hat{P} + (1 - P)(1 - \hat{P})$$

Donde P corresponde al porcentaje de positividad de las observaciones y \hat{P} corresponde al porcentaje de positividad de las proyecciones. La varianza del ratio SRI se define como:

$$VAR[SRI] = n^{-2} \left[n(2 \cdot \hat{P} - 1)^2 P(1 - P) + n(2 \cdot P - 1)^2 \hat{P}(1 - \hat{P}) + 4 \cdot P \cdot \hat{P}(1 - P)(1 - \hat{P}) \right]$$

Por su parte, la varianza del ratio de éxito (SR) se define como:

$$VAR[SR] = n^{-1} \cdot SRI(1 - SRI)$$

Finalmente, el test DA de Pesaran & Timmermann (1992) está dado por:

$$DA = (VAR[SR] - VAR[SRI])^{-1/2} (SR - SRI)$$

Como dicho test sigue una distribución normal estándar, los valores de rechazo de la hipótesis nula serán el -1,96 y el 1,964. Es decir; si el valor DA está entre los valores de rechazo, no se rechaza la hipótesis nula de que las variaciones observadas están independientemente distribuidas de las variaciones proyectadas.

De esto último se entiende que se busca rechazar la hipótesis nula; es decir, que el valor DA no se encuentre entre los valores críticos mencionados y que por ende si exista capacidad predictiva.

⁴ Al 5% de nivel de significancia.