

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS

ESCUELA DE ECONOMÍA Y ADMINISTRACIÓN

“Administración de carteras con redes neuronales mediante metodología Rolling”

Seminario para optar al título de ingeniero comercial

Profesor guía: Antonino Parisi Fernández

Autor: Marcelo Gutiérrez Márquez

Santiago de Chile

Diciembre 2004

A mi familia por su amor y apoyo constante

*A María Inés Campo por su compañía,
apoyo y ánimo incondicionales*

*A Antonino Parisi, por confiar en mí para este tema,
por su guía y su optimismo constante*

*A David Díaz, Julio Rebolledo Y Edinson Cornejo
por su excelente disposición y generosa
y desinteresada colaboración*

*Y a mi abuelo, por inculcarme el amor por la
Universidad de Chile y por el conocimiento*

“Administración de carteras con redes neuronales mediante metodología Rolling”

Marcelo Gutiérrez Márquez

Profesor Guía: Antonino Parisi Fernández

Abstracto

Este trabajo persigue evaluar la rentabilidad que habría obtenido un inversionista que hubiese seguido las recomendaciones de Redes Neuronales Artificiales para la conformación semanal de sus carteras durante casi 4 años. Lo que se busca es ver no sólo si la proyección de las RNA constituyen una mejor predicción del retorno de las acciones que el promedio histórico de estos, si no que además busca ver si al usar una metodología Rolling para las predicciones se puede lograr un mejor desempeño de las RNA y por ende una mayor rentabilidad de la conformación de carteras. Para esto se usa una red Ward que se vuelve a entrenar con los nuevos datos cada vez que se quiere proyectar una nueva semana (Método Rolling). En este estudio se usaron datos de 29 de las 30 acciones del Dow Jones Industrial Average para el período comprendido entre el 4 de febrero de 1994 y el 10 de septiembre del 2004. Como Benchmark para el método propuesto se usó la metodología tradicional de conformación de carteras tomando el coeficiente Beta de correlación con el mercado como medida del riesgo y el promedio de los retornos históricos como medida del retorno esperado, este método se denominará como portfolios por Betas. En este trabajo se encontró que la red Ward tenía excelente capacidad predictiva para el signo que seguirá el cambio del precio de una acción. También se mostró que cuando se permite venta corta la estrategia por RNA obtiene mayor rentabilidad acumulada en promedio que la estrategia por Betas. En cambio, cuando no se permite la venta corta los portfolios por Betas superan a los de RNA en este mismo indicador.

INDICE

Parte I

I.1 Introducción.....	5
I.2 Marco Teórico.....	6
I.2.1 Redes Neuronales Artificiales (RNA).....	9
I.2.1.a RNA: Sistemas paralelos, distribuidos y adaptativos.....	10
I.2.1.b LA Neurona.....	11
I.2.1.c Funciones de activación.....	14
I.2.1.d Arquitectura de RNA.....	16
I.2.1.e Proceso de aprendizaje.....	18
I.2.1.f Generalización v/s memorización.....	19
I.2.1.g Síntesis sobre redes neuronales.....	20

Parte II

II.1 Metodología.....	22
II.1.1 Método Rolling.....	22
II.1.2 RNA Utilizada.....	23
II.1.3 Modelo para la rentabilidad.....	24
II.1.4 Conformación de portfolios.....	24
II.1.4.a Según los betas.....	24
II.1.4.b Según RNA.....	26
II.1.5 Resultados comparables.....	28
II.1.6 Validación y capacidad predictiva de los modelos.....	29
II.1.7 Datos.....	31

Parte III

III.1 Resultados.....	33
III.1.1 Capacidad predictiva.....	33
III.1.2 Portfolios.....	34

Parte IV

IV.1 Conclusiones.....	37
Anexo 1: “Evolución del Dow Jones diciembre 2000 a diciembre 2004”	39
Referencias	41

Parte I

I.1 INTRODUCCIÓN

Hoy en día las redes neuronales son aplicadas en muchas áreas como la geología, física o matemáticas. Incluso son utilizadas por las isapres para detectar licencias falsas. En el área de las finanzas se han aplicado para predecir precios de commodities como el oro, para estimar la probabilidad de quiebra y para evaluar la calidad crediticia, entre muchas otras aplicaciones.

Esta tesis pretende evaluar como se habría desempeñado un administrador de carteras de inversión o un inversionista que hubiese seguido las recomendaciones de compra y venta dadas por las predicciones de una red neuronal artificial, reconfigurando su portfolio de forma semanal, durante 190 semanas consecutivas. Como benchmark utilizaremos las estrategias de “buy and hold” y portfolios conformados utilizando el beta como medida de riesgo.

Para la conformación de las carteras de cada semana se usará un proceso “Rolling” en el que los datos más nuevos irán siendo agregados y los más antiguos se irán eliminando de las series usadas en el estudio.

I.2 MARCO TEÓRICO

Al conformar un portfolio de inversión, la metodología común es la de maximizar el retorno dado un riesgo requerido o minimizar el riesgo para un nivel dado de retorno exigido. A partir de esto se decide la proporción óptima a invertir en cada activo disponible. Esta estrategia se desprende de la teoría de carteras de Markovitz, propuesta en 1952. En este caso maximizaremos el retorno del portfolio para un nivel de riesgo máximo que estaremos dispuestos a soportar.

Las 2 formas tradicionales y más comunes de conformar portfolios utilizan la esperanza del retorno de cada acción como el más probable que puedan lograr en el próximo período, o sea que usan el promedio de los retornos pasados como medida del retorno esperado y es este el que usan para componer el máximo retorno para el portfolio.

Esta metodología tradicional de conformación de carteras, tiene asociado el crucial principio de la diversificación, ya que reparte la riqueza entre varios activos, lo que ayuda a disminuir el riesgo no sistemático.

Sin embargo estos métodos difieren en su medida del riesgo. Como sabemos, el riesgo es básicamente la incapacidad de predecir lo que pasará en el futuro con el retorno de las acciones. Así, el primer método lo representa a través de la varianza histórica de las acciones, fundamentado en la idea de que esta nos dirá cuánto podría llegar a alejarse el retorno efectivo del retorno esperado, o sea cuál es la volatilidad del retorno de cada acción. El segundo método usa el coeficiente Beta de correlación de la acción con el portfolio de mercado, o sea toma como medida de riesgo un coeficiente que indica como se mueve el retorno del activo cuando se mueve el retorno de mercado. Nuevamente se usan datos históricos para calcular este coeficiente beta.

Pero estos métodos asumen que no es posible predecir el camino que seguirán los retornos de los activos que conformarán la cartera, o sea asumen que estos siguen un proceso de “random walk” o “camino aleatorio”, por lo que no tratan de predecir los retornos, si no que ven como se ha comportado en el pasado y según eso esperan poder estimar lo que ocurrirá en el futuro. Pero esto es hecho sabiendo que no necesariamente los rendimientos pasados representan lo que ocurrirá en el futuro.

Este concepto es respaldado por Eugene Fama en 1970, quién publica la llamada “*Hipótesis de Mercados Eficientes*”, en la que establece que los retornos de los activos financieros no son predecibles, ya que si el mercado es “*eficiente*”, entonces el precio de un activo sería capaz de incorporar toda la información disponible y necesaria hasta el momento en que se valore ese activo. Esto, sumado a la suposición de que los agentes son racionales, implica que no es posible para un participante de este mercado obtener retornos anormales, o sea que nadie sería capaz de “ganarle” al mercado a menos que logren adivinar el futuro o predecir la información que aún no existe, ya que es sólo este tipo de información la que podría producir un cambio en el precio. Pero la nueva información no puede estar relacionada con la antigua (de serlo, ya habría sido considerada e incorporada al precio del activo) lo que la hace impredecible.

Pero Fama hace un supuesto fundamental que no se cumple en la realidad y es el de que los costos de transacción e información sean iguales a cero. Este solo supuesto ya nos permite suponer que en el mundo real sí se pueden obtener rentas superiores a las de mercado.

Muchos estudios contradicen o por lo menos reinterpretan los conceptos anteriormente expuestos. Por ejemplo, Campbell, Lo y MacKinlay en 1997 y Gavridis en 1998, junto a varios otros estudios establecieron que el mercado accionario puede tener una parte predecible, sin dejar de ser eficiente. Esto se

debería, según ellos, a que habría un componente de largo plazo del mercado que sería posible modelar y proyectar.

De hecho, antes de juntarse con Campbell; Lo y MacKinlay ya habían escrito en 1988 su paper llamado “Los precios accionarios no siguen un *random walk*: evidencia sacada de una prueba simple de especificación”. En el que, como lo dice su nombre, encontraron evidencia de que los precios accionarios no siguen necesariamente un “random walk” y de que serían predecibles en cierto grado, basados en que identificaron una correlación sistemática positiva entre los retornos semanales de los principales mercados mundiales.

Muchos otros estudios han encontrado evidencia sobre la predictibilidad de los retornos tanto a corto como largo plazo. También Ferson y Harvey en 1991 descubrieron que muchas variables que sirven para explicar la rentabilidad de los activos son predecibles. Lo que haría que los retornos de las acciones sean predecibles, por lo menos de forma indirecta. Además, Leung, Daouk y Cheng en el año 2000 concluyen que la predicción de los retornos accionarios es un hecho real, haciendo que la pregunta ya no sea si se puede o no predecir los retornos, si no que como se debe usar la información de forma óptima para poder predecirlos. En este mismo estudio se demostró que se podían obtener mejores resultados y por ende mayores utilidades al intentar predecir el signo de los cambios de los precios de los activos en vez de tratar de predecir el nivel exacto que alcanzarán.

Así, en todos estos últimos estudios y muchos otros podemos fundamentar la idea de querer predecir lo que ocurrirá en el futuro con los retornos de las acciones y según eso conformar nuestros portafolios de forma semanal (en este caso), ya sea a partir de análisis técnico o el análisis fundamental. Esto, claro, reconociendo que las redes neuronales o cualquier otro método matemático que usemos para predecir lo que ocurrirá con los retornos no podrán conocer lo que ocurre en el “exterior” o sea no podrán considerar datos como la ocurrencia de un desastre natural o que se lleve a cabo un atentado terrorista, por lo que siempre el

criterio del analista seguirá jugando un papel importante a la hora de tomar decisiones de inversión y conformación de carteras.

I.2.1 REDES NEURONALES ARTIFICIALES (RNA)

Las Redes Neuronales Artificiales (RNA) son sistemas computacionales que intentan imitar el cerebro humano, hasta en su capacidad de “aprendizaje”. En otras palabras, uno puede entrenar o “educar” a una red neuronal mostrándole todas las variables que influyen en un problema y sus respectivos resultados y la red neuronal irá probando distintas relaciones matemáticas o ponderadores para cada variable dependiente, hasta encontrar la relación que mejor se ajuste o que minimice el error de la predicción. Esto se puede lograr gracias a que a la red se le muestran no solo las variables independientes sino que también las dependientes que han ocurrido en el pasado. Después se puede probar lo que la red ha “aprendido” mediante un grupo de datos de prueba, donde nuevamente se le muestran tanto las variables dependientes como las independientes que han ocurrido en el pasado, pero esta vez la red no va adaptando sus ecuaciones, sino que aplica la ecuación o relación a la que llegó en la etapa de aprendizaje y compara el resultado que obtiene con el que se dio en la realidad, pudiendo medir de esta forma que tanto se alejan en promedio las predicciones de la red de lo que ocurre en la realidad. Si el analista queda satisfecho con la calidad de ajuste que ha logrado la red, puede proceder a aplicar la red neuronal a otro grupo de datos que nunca ha visto, solo que esta vez a la red neuronal se le entregarán únicamente las variables independientes, a partir de las cuales deberá predecir las dependientes.

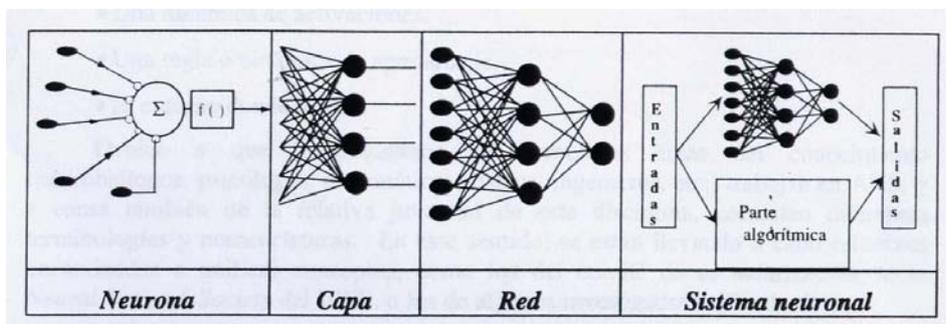


Fig. 1: Estructura jerárquica de un sistema basado en RNA
Fuente: “Redes neuronales y sistemas borrosos”. Del Brío y Sanz

Para el desarrollo de esta tarea la red neuronal está compuesta de varias capas, las que a su vez están compuestas de varias neuronas. Cada neurona se “reparte” una porción del problema con lo que son capaces de trabajar al mismo tiempo con la porción que le corresponda. Además, las neuronas de una capa pueden generar inputs para las neuronas que vienen en las capas posteriores. Esto dependerá de la cantidad de capas que posea la red neuronal.

Básicamente las redes neuronales tienen una capa de entrada con una cantidad de neuronas igual a la cantidad de inputs del problema, luego una o más capas ocultas que poseen la cantidad de neuronas que el analista estime conveniente y finalmente hay una capa de salida que posee solo una neurona. La capa de entrada recibe los datos y los reparte entre las neuronas de la primera (y a veces única) capa oculta, sin mayor procesamiento de los datos más que la posible aplicación de una función escalar. Luego los datos entran a las capas ocultas donde son procesados y se les aplica una función de activación, definida por la persona que aplica la red para cada capa oculta. Después de pasar por las capas ocultas, todos los datos llegan a la capa de salida, donde se procesan los datos y se entrega como output la respuesta buscada.

I.2.1.a RNA: SISTEMAS PARALELOS, DISTRIBUIDOS Y ADAPTATIVOS¹

Existen tres características fundamentales del sistema nervioso humano que las RNAs buscan imitar: Poder formular varios cálculos a la vez de forma **paralela**, poder **distribuir** la memoria disponible y ser capaz de **adaptarse** al entorno.

El **Procesamiento paralelo** es esencial si se quiere imitar al sistema nervioso humano, ya que el hombre es capaz de procesar varios datos al mismo tiempo, lo que le permite entender las cosas de forma mucho más rápida que si

¹ Extracto del capítulo 1 del libro “*Redes neuronales y sistemas borrosos*” de Martín del Brío y Alfredo Sanz Molina

tuviera que ir interpretando lo que sucede en su entorno un paso a la vez, por ejemplo, las neuronas de la visión (los conos y bastones de la retina) se “reparten” la visión, encargándose cada una de una porción pequeña de lo que el ojo alcanza a ver. De la misma forma la red neuronal reparte distintas “porciones” del problema entre todas sus neuronas o *procesadores elementales*, con lo que logra procesar los datos y encontrar una solución de forma más rápida.

Otro concepto importante es el de la **memoria distribuida**. En un computador la información se almacena en posiciones o lugares físicamente bien definidos (o sea se le asigna un espacio en el disco duro o en la memoria temporal). Sin embargo, en los sistemas neuronales la información se distribuye según las conexiones de la red (o sinapsis, que es como se denominan las conexiones entre neuronas). De esta forma, si una conexión se daña solo se perderá una pequeña fracción de la información. Además, los sistemas neuronales biológicos son redundantes, de forma que muchas neuronas y sinapsis pueden realizar un papel similar; o sea, el sistema tolera fallas.

La tercera característica primordial es la de la **adaptabilidad**. Las RNA poseen una gran variedad de mecanismos para adaptarse a su entorno o para “aprender” según las influencias que recibe del exterior. La forma más común que tienen las RNA para adaptarse es mediante la modificación de sus sinápsis (conexiones entre neuronas), lo que les permite aprender de la experiencia. Esto les permite generalizar a partir de casos particulares.

Así, podemos hablar de las RNA como sistemas **Paralelos, distribuidos y adaptativos**.

I.2.1.b LA NEURONA

Como ya se puede desprender de lo leído anteriormente, la neurona es la estructura básica de las redes neuronales. Éstas, intentan imitar a las que

tenemos en el cerebro, recibiendo impulsos desde varias fuentes, procesando los impulsos y retransmitiendo sus propios impulsos a otras neuronas o a los receptores finales de la información (como los músculos por ejemplo). De esta misma forma, las neuronas que componen las RNA pueden recibir información desde varias fuentes a la vez, aplicarles una función matemática y posteriormente traspasar sus resultados a otras neuronas o al analista que está esperando los resultados.

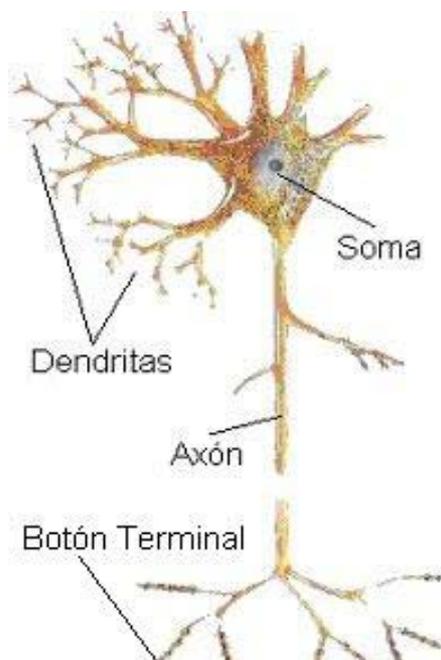


Fig.2: Neurona biológica típica
Fuente: www.cienciateca.com

terminales, donde el impulso puede ser traspasado a las dendritas de otra neurona o directamente a un músculo si se trataba de una neurona motora. Así los impulsos nerviosos viajan a la velocidad de la luz desde el lugar donde se generan

En la **Figura 2** se puede apreciar la estructura de una neurona biológica típica. En este diagrama se aprecian sus partes funcionales básicas. La neurona recibe impulsos eléctricos desde otras neuronas o directamente desde el exterior a través de sus dendritas por un proceso denominado "sinápsis". Gracias a que posee gran cantidad de dendritas puede recibir información de varias fuentes a la vez. Luego los impulsos llegan al soma, que es el lugar donde se procesan estos impulsos y se sintetiza la información recibida en un solo impulso que viajará por el axón hasta los botones

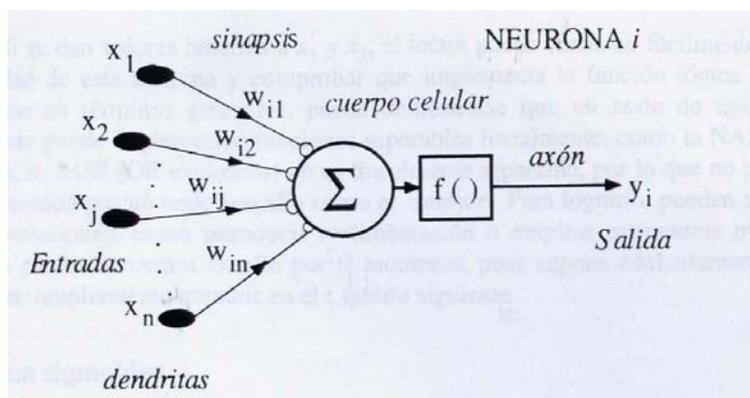


Fig. 3: Neurona de una RNA
Fuente: "Redes neuronales y sistemas borrosos". Del Brío y Sanz

hasta su destino a través de miles de neuronas, cada una con una función específica.

En la **Figura 3** se puede ver un diagrama básico de una neurona o “Procesador Elemental” (PE) de una RNA. Ahora es fácil apreciar la similitud entre las neuronas biológicas y los PE de las RNA. Tenemos que la neurona recibe información desde distintas fuentes, esta vez, la información consistirá de datos provenientes de otras neuronas. A cada dato le asignará un peso (W_{ij}) que podrá tomar valores entre 1 y -1, este dependerá de la intensidad de interacción entre la neurona que genera el dato y la que lo recibe, el valor del peso dependerá de:

- Si la neurona precedente (neurona i) incentiva la activación de la neurona que recibe la información (neurona j), entonces la conexión es excitatoria ($w_{ij} > 0$).
- Si la neurona precedente (neurona i) desincentiva la activación de la neurona que recibe la información (neurona j), entonces la conexión es inhibitoria ($w_{ij} < 0$).
- Si la neurona precedente (neurona i) no ejerce efecto alguno sobre la activación de la neurona que recibe la información (neurona j), entonces la conexión es llamada “inactiva” ($w_{ij} = 0$).

Luego, mediante la **regla de propagación** que haya definido previamente el usuario de la RNA. La neurona tomará todos los datos recibidos y les calculará un valor neto (llamado entrada neta). La forma más común que toma esta regla es la de una suma ponderada de los datos por sus pesos respectivos, con lo que queda:

$$\sum_j w_{ij} x_j$$

Donde X_j es el dato que proviene de la neurona “j”. Luego a esta suma ponderada se le aplicará la función de activación (“f()” en la **figura 3**) que el

programador le haya asignado a la capa a que pertenece esa neurona. Así la neurona generará un output de:

$$y_i(t) = f_i \left(\sum_j w_{ij} x_j \right)$$

El que servirá de entrada para una o más neuronas posteriores o si la neurona era parte de la capa de salida, entregará el resultado de la red.

I.2.1.c FUNCIONES DE ACTIVACIÓN

Las funciones de activación son otra parte fundamental de las redes neuronales, estas constituyen la forma en que las neuronas manipularán los datos que reciban para generar la información que le transmitirán a la siguiente capa de neuronas o al analista si forman parte de la capa de salida. Las funciones de activación son definidas por el usuario de la red y se le asigna a cada capa de neuronas una función de activación.

Existen infinitos tipos de funciones de activación posibles, las que se usen dependerán de cómo quiera el usuario que la red manipule los datos o de que importancia le dará a distintos tipos de éstos. Por ejemplo, si se usa una función identidad o lineal, donde $y = x$ (o $F(x) = x$), el analista le estaría dando la misma importancia a todos los datos. Otro tipo de función podría ser de tipo “escalón” en la que la neurona no generará información a menos que la entrada neta de información alcance un cierto “umbral”.

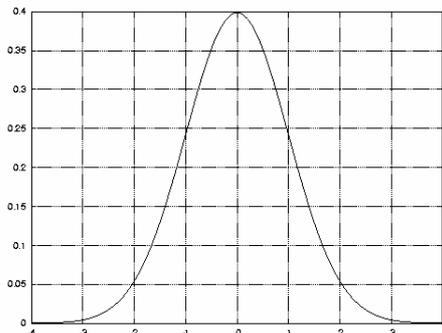


Fig. 4: Función de distribución Gaussiana

Para fines de esta investigación nos interesarán la función *Gaussiana* y *complemento de Gaussiana*, ya que serán estas dos las que serán usadas.

La función **Gaussiana**, no es más que la

muy conocida función de distribución normal o también llamada “campana de Gauss” que podemos ver en la **figura 4**. Esta función es monótona creciente y se usa cuando se cree que la información del centro o cercana a la media es la más relevante, por lo que se concentra en detectar diferentes características en este rango de datos. Además entrega datos de salida dentro del rango (0,1). La fórmula de la función de distribución Gaussiana es:

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

La función **Complemento de Gaussiana** es la inversa de la función Gaussiana, o sea es una campana de Gauss “de cabeza”, como muestra la **figura 5**. Esta función le asigna más importancia a los datos de las colas o de los extremos de la distribución, como se puede ver en el gráfico. Su expresión matemática es muy similar a la función Gaussiana y es:

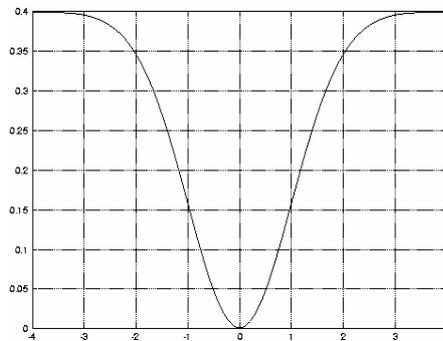


Fig. 5: Función de distribución complemento de Gaussiana

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

Al usar estas dos funciones en dos slabs² distintos dentro de la capa oculta, se logra que la capa de salida tenga dos perspectivas de los datos. Al combinar los dos puntos de vista de las características de los datos, se puede lograr una mejor predicción.

² Un “Slab” es un conjunto de neuronas que pertenecen a la misma capa y usan la misma función de activación. Puede existir más de un slab dentro de cada capa. De hecho en esta tesis la capa oculta poseerá dos slabs, como se verá más adelante.

Hay que decir que las neuronas de la capa de entrada también tienen funciones de activación, pero éstas por lo general son transformaciones lineales. Lo que buscan estas funciones es transformar los datos de entrada a un formato más “amigable”, dejando todos los datos de entrada en la misma escala para que la red neuronal trabaje con ellos, llevándolos todos a la misma escala, haciéndolos comparables. En otras palabras “traduce” los datos a un lenguaje más fácil de entender para la red neuronal. Después, cuando los datos finales salen de la capa de salida, son “des - traducidos” aplicándoles la función inversa de la función de activación de la capa de entrada.

I.2.1.d ARQUITECTURA DE RNA³

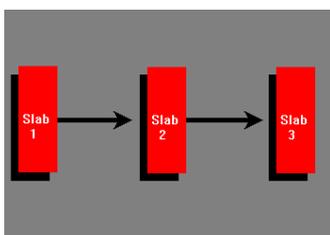


Fig. 6: Red de 3 capas con arquitectura estándar

La arquitectura de una RNA constituye la forma en que están distribuidas las neuronas, grupos de neuronas (slabs), capas y como están conectadas e interactúan entre sí. También definirán como fluye la información a través de la red.

La forma en que están conectadas las neuronas determinará la forma en que se comporte la red, por lo que la arquitectura es una característica fundamental de la RNA que se use en distintos estudios.

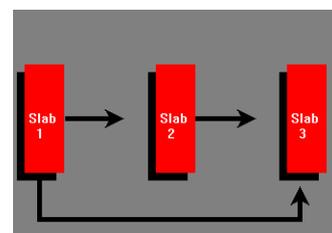


Fig. 7: Red de 3 capas con conexión jump

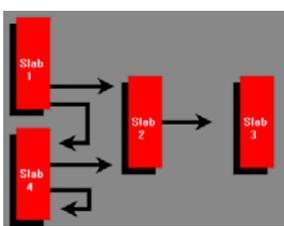


Fig. 8: Red Jordan – Elman de 3 capas

Existen muchos tipos distintos de arquitecturas para redes neuronales, pero entre las más populares se encuentran las redes tipo estándar como la de la **figura 6** que posee 3 capas. Otro tipo popular de RNA son las que tienen conexiones de tipo “jump” (Salto en inglés), como la

³ Fuente de las ilustraciones de las distintas arquitecturas: Software de redes neuronales “Neuroshell 2”

que aparece en la **figura 7**, del esquema se puede ver que se les llama conexiones jump debido a que hay conexiones entre capas que no son sucesivas en la red. Un tercer tipo de redes más comúnmente usadas son las de *Jordan – Elman* ilustrada en la **figura 8** una particularidad de estas redes es que los grupos de neuronas se pueden “realimentar” con su propia información o con la información de una capa paralela.

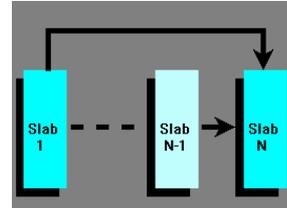


Fig. 9: Red polinomial GMDH

Se puede ver que este tipo de redes la capa de entrada tiene 2 slabs de neuronas, o sea que hay un slab extra. Esto le permite a la red ver el conocimiento anterior que tenía sobre las entradas anteriores. El cuarto tipo de arquitectura que deseo mencionar es la red polinomial GMDH (Group Method of Data Handling) o la red polinomial de Método de Manejo Grupal de Datos que se puede ver en la **figura 9**. Esta red va construyendo capas sucesivas con conexiones complejas que constituyen los términos de un polinomio. Estos términos del polinomio son creados a partir de regresiones lineales y no – lineales, que se van generando en cada capa a partir de los resultados obtenidos de las capas anteriores.

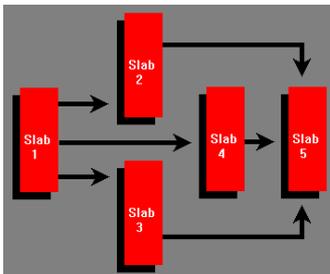


Fig. 10: Red Ward de 3 capas

El último tipo de red que nos interesa es la red “*Ward*”, que será la red usada en este trabajo. Como se aprecia en la **figura 10**, la red Ward permite tener más de un slab de neuronas en las capas ocultas. Con esto se le pueden aplicar distintas funciones de activación a cada slab de la capa oculta, lo que permite detectar diferentes características en los patrones procesados a través de la red. Así se pueden aplicar las funciones de activación Gaussiana y complemento de Gaussiana, que mencionamos anteriormente, dentro de la misma capa oculta, logrando que la capa de salida tenga diferentes puntos de vista de los datos, como también se dijo con anterioridad.

Como se dijo anteriormente existen muchos tipos de arquitecturas de redes y muchas variaciones posibles para las redes acá presentadas.

I.2.1.e PROCESO DE APRENDIZAJE

Como ya dijimos, las redes neuronales “aprenden” relaciones entre distintas variables a partir de la identificación de patrones observables en datos pasados.

La forma más común de “aprendizaje” y también de adaptación, que tienen las redes neuronales es a través de la modificación de los pesos (W_{ij}) que cada neurona le asigna a la información que recibe de las neuronas que la preceden. Así, las RNA van modificando los pesos en la etapa de entrenamiento para ir reduciendo el error que cometen al tratar de predecir las variables independientes. Como se dijo antes, en la etapa de entrenamiento las redes neuronales van calculando sus predicciones de las variables requeridas por el usuario y cada vez que llegan a un valor lo comparan con el valor real que tuvo la variable que se intentó predecir, así la red ve que tanto se equivocó y modifica los pesos sinápticos (W_{ij}) para tratar de disminuir ese error. Para esto la red neuronal puede partir con pesos iguales a cero ($W_{ij} = 0$) o asignar pesos de forma aleatoria antes de empezar a “aprender”, así tiene un punto desde donde empezar a adaptar los pesos. Una forma muy común de ir modificando los W_{ij} es mediante una distribución proporcional del error cometido en la predicción entre todas las neuronas según el aporte de cada una al resultado final, o sea, mientras más aportó la neurona o conexión al resultado, mayor parte del error se le asignará a ella para modificar sus pesos.

En su libro “*Redes neuronales y sistemas borrosos*” Bonifacio Martín del Brío y Alfredo Sanz Molina definen el proceso de aprendizaje como:

“...proceso por el que se produce el ajuste de los parámetros libres de la red a partir de un proceso de estimulación por el entorno que rodea la red. El tipo

de aprendizaje vendrá determinado por la forma en la que dichos parámetros son adaptados. En la mayor parte de las ocasiones el aprendizaje consiste simplemente en determinar un conjunto de pesos sinápticos que permita a la red realizare correctamente el tipo de proceso deseado.”

Otra forma de aprender o adaptarse de las redes neuronales es a través de la creación o destrucción de neuronas, en el que la red sufre un cambio de su propia arquitectura.

I.2.1.f GENERALIZACIÓN V/S MEMORIZACIÓN

La gracia del aprendizaje de las RNA es que puedan generalizar a partir de ejemplos, lo que constituye una de las características fundamentales de las redes neuronales. Esta característica permite que las RNA puedan generar datos confiables con datos que no han visto con anterioridad a partir de lo que han aprendido de la información pasada que se les entregó en el período de aprendizaje.

Pero al esforzarse por reducir el error (medido por el Error Cuadrático Medio o ECM) que comete en la etapa de aprendizaje, puede que la red memorice los patrones de la fase de entrenamiento, con lo que la red serviría solo para

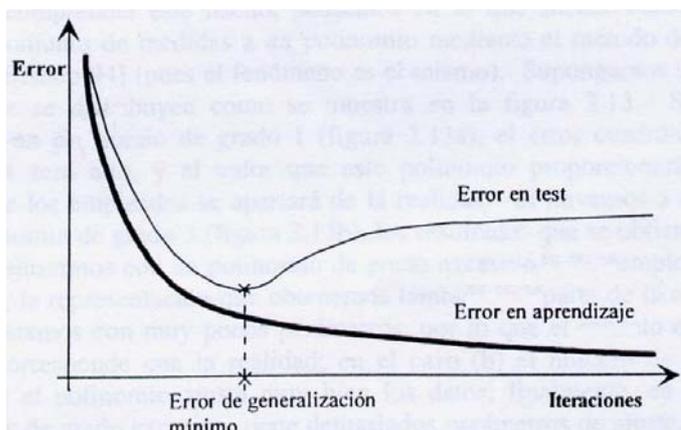


Fig. 11: *Evolución del error de aprendizaje y de generalización*
Fuente: “Redes neuronales y sistemas borrosos”. Del Brío y Sanz

predecir valores para ese conjunto de datos y no para datos que no ha visto nunca. Por ejemplo, si dejamos que la red haga todas las iteraciones que sean necesarias hasta obtener un ECM = 0 en la etapa de entrenamiento, la red se ajustará perfectamente a estos datos, pero quizás cuando le

presentemos datos nuevos, su ajuste será pésimo. De hecho ya se han llevado a cabo experimentos para probar esto y se ha llegado a determinar que a medida que aumentan las iteraciones hechas en la etapa de entrenamiento, el error de la fase aprendizaje y el error de la etapa de prueba (o test) disminuyen juntos en una primera etapa, pero después de cierto punto el error de la etapa de prueba empieza a crecer nuevamente en circunstancias que el ECM para el conjunto de entrenamiento sigue disminuyendo, como muestra el gráfico de la **figura 11**. Es por esto que por lo general se usa un conjunto de datos para entrenar la red y otro conjunto distinto para probar que tan buen ajuste posee. Así, el ECM de la red usando los datos de prueba es un buen indicador de la capacidad de generalización que posee la red o de que tanto memorizó los datos de entrenamiento.

I.2.1.g SÍNTESIS SOBRE REDES NEURONALES

Así, podemos decir que las redes neuronales son sistemas **paralelos**, **distribuidos** y **adptativos** que buscan imitar al cerebro en sus características fundamentales, sobre todo en el aspecto de lograr aprender relaciones sobre distintas variables a partir de información histórica. Para lograr esto, las RNA se componen de⁴:

- Un conjunto de procesadores elementales o neurona artificiales. (Que procesarán la información)
- Un patrón de conectividad o arquitectura. (Que determinará el orden y la forma en que se procese la información)
- Una dinámica de activaciones. (Que determinará como se tratará la información y que importancia se le dará a cada tipo)
- Una regla o dinámica de aprendizaje. (Que determinará que tan buena es la red para generalizar a partir de ejemplos pasados)

⁴ Según D.E. Rumelhart y J.L. McClelland del Parallel Distributed Processing Research Group de la Universidad de California en San Diego.

- El entorno donde opera. (Que influenciará la forma en que la red se adapte o aprenda de los datos que se le entreguen).

Esta es una “pincelada” de lo que son las redes neuronales. Esta reseña está enfocada principalmente a este trabajo de investigación y sus necesidades. Para una profundización del tema recomiendo el libro “Redes neuronales y sistemas borrosos” de Bonifacio Martín del Brío y Alfredo Sanz Molina, de la editorial RA – MA

Parte II

II.1 METODOLOGÍA

Lo que persigue esta tesis es evaluar como le habría ido a un inversionista si hubiese conformado su portfolio de acuerdo a las recomendaciones de una red neuronal durante 190 semanas (3 años y 8 meses aproximadamente). Para evaluarlo se tomaron datos pasados con los que se fueron conformando carteras semanalmente, según lo recomendado por la RNA y se calculó la rentabilidad que habría obtenido cada cartera según lo que ocurrió en la realidad. Estos resultado se compararon tanto con portfolios conformados según los Betas (coeficientes de correlación con el mercado) de cada acción y con una estrategia ingenua o de "Buy & Hold". Además, para efectos de la compra y venta de acciones se consideró un costo US\$ 10 por transacción, para hacer más realista el ejercicio.

Para el desarrollo de esta investigación se usó el software Microsoft Office Excel, en el que se programó la red neuronal mediante el uso del software, también de Microsoft, Visual Basic. Además se usó el complemento de Excel llamado "Super Solver" que es una versión más poderosa y moderna del típico Solver que viene junto con Excel.

II.1.1 MÉTODO ROLLING

Para ir sacando las predicciones de la red neuronal se usó el método de "Rolling" que consiste en que cada vez que la red neuronal debe hacer una predicción para la siguiente semana, esta se vuelve a entrenar. Pero esta vez se quita del conjunto de entrenamiento la semana más alejada en el tiempo y se agrega una semana más nueva que todas las ya existentes en el conjunto. Así, la red recalcula los pesos sinápticos para cada semana que va a predecir.

Este método se usó basado en las recomendaciones que se hacen en el paper Parisi, Parisi y Guerrero (2003). Los que concluyeron que los pesos

sinápticos, (w_{ij}) por los que se pondera cada “trozo” de información que entra a cada neurona, no eran estables a través del tiempo.

II.1.2 RNA UTILIZADA

Para este estudio se usó una red Ward, siguiendo nuevamente las recomendaciones de Parisi, Parisi y Guerrero (2003), ya que en ese paper encontraron que esta era la mejor arquitectura para este tipo de predicción.

Como ya se dijo anteriormente, la red Ward permite que se usen distintas funciones de activación en cada slab. Así, para esta investigación se usó la función Gaussiana en un slab de la capa oculta y una función Complemento de Gaussiana en la otra. Nuevamente, como ya se mencionó con anterioridad,

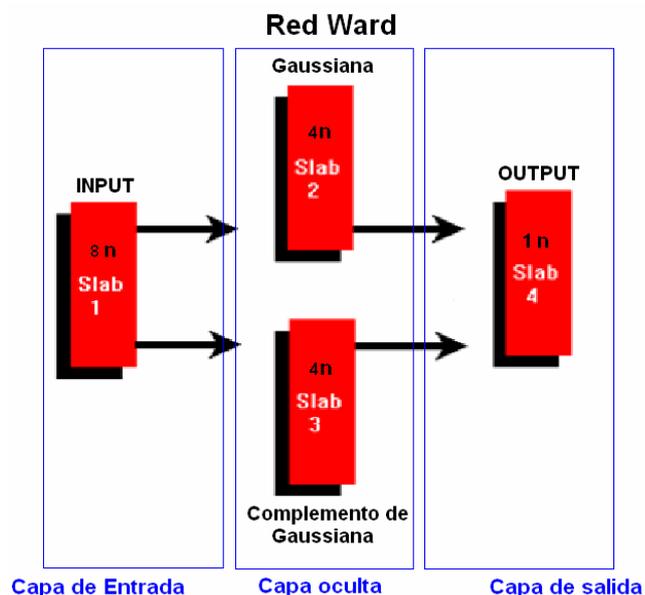


Fig. 12: Esquema de la RNA usada en este trabajo

esto permite que la red tenga dos puntos de vista de la misma información, con lo que se puede lograr un resultado más adecuado a la realidad. Como muestra la **figura 12**, la red usada posee 3 capas. En su capa de entrada posee 8 neuronas (una por cada variable del modelo usado), la capa oculta posee 2 slabs de 4 neuronas cada uno y la capa de salida posee una neurona.

II.1.3 MODELO PARA LA RENTABILIDAD

Para modelar la rentabilidad de las acciones se usó una ecuación en que la rentabilidad nominal (medida en centavos de dólar) de cada acción para la próxima semana depende de la variación que tuvo el Dow Jones hace una, dos, tres y cinco semanas y de las variaciones del precio de la acción hace una, dos, tres y cuatro semanas atrás. Con esto tenemos que el modelo para la rentabilidad de la próxima semana es:

$$Var(x) = Var(VarDow_{-1}, VarDow_{-2}, VarDow_{-3}, VarDow_{-5}, VarX_{-1}, VarX_{-2}, VarX_{-3}, VarX_{-4})$$

Donde "X" corresponde a la acción cuya variación se busca predecir y los subíndices indican cuantas semanas antes de la predicción fue tomado el dato.

II.1.4 CONFORMACIÓN DE PORTFOLIOS

Como ya se dijo antes, compararemos los portfolios conformados por redes neuronales con los conformados a partir de los betas de cada acción. Por lo que a continuación se explican ambos métodos de conformación de carteras.

II.1.4.a SEGÚN LOS BETAS

El riesgo de un portfolio de acciones dependerá del riesgo de los activos que lo compongan y en que proporción lo hacen. De la misma forma el retorno esperado de la cartera dependerá del retorno esperado de sus componentes.

Así, el beta de una cartera es el promedio ponderado de los betas de las acciones por la proporción de la cartera invertida en cada una. Por ende, una vez que se tiene el Beta histórico para cada acción, el Beta del portfolio sería:

$$\beta_{portfolio} = \sum_{i=1}^n \beta_i * \lambda_i$$

$$Tal \quad que : \sum_{i=1}^n \lambda_i = 1$$

Donde β_i corresponde al Beta de la acción "i" y λ_i representa la proporción de la cartera invertida en el activo "i". A su vez, la rentabilidad esperada de la cartera corresponderá a la suma de las rentabilidades esperadas de cada acción por la proporción invertida en cada una de ellas. Nuevamente un promedio ponderado. Así, la rentabilidad esperada del portfolio es:

$$E(R_{portfolio}) = \sum_{i=1}^n E(R_i) \lambda_i$$

$$Tal \quad que : \sum_{i=1}^n \lambda_i = 1$$

Donde $E(R_i)$ representa la rentabilidad esperada del activo "i", la que es calculada sacando el promedio de los retornos históricos más recientes⁵. Aquí se puede ver una de las fallas de la estrategia de Markovitz y es la falta de una medida adecuada para el retorno esperado, esto es lo que busca mejorar esta tesis mediante el uso de redes neuronales.

Así, a partir de estos datos se busca maximizar el retorno esperado del portfolio sujeto a un nivel dado de riesgo máximo que está dispuesto a soportar el inversionista, mediante el uso de una aproximación cuadrática. Así el problema de maximización es:

⁵ La cantidad de semanas hacia atrás que se use dependerá del criterio del analista

$$\text{Máx } E(R_{\text{portfolio}}) \text{ sujeto a: } \beta_{\text{portfolio}} = \beta^*$$

Donde β^* es el Beta máximo que el inversionista está dispuesto a soportar en su cartera.

Para resolver esta maximización se usa el programa Super Solver, que es la versión mejorada de un complemento de Excel, que permite que el computador haga estas maximizaciones iterando hasta encontrar la mejor solución.

II.1.4.b SEGÚN RNA

Al conformar portfolios a partir de RNA se utilizará la misma metodología básica de maximizar el retorno de la cartera, sujeto a un nivel de riesgo máximo. Pero por este método se usan distintas medidas tanto para el retorno esperado como para el riesgo.

Primero, el retorno esperado de cada acción ya no corresponderá a la media histórica de este, si no que al retorno que la red neuronal prediga para la acción cada semana. Pero el retorno de la cartera seguirá siendo la suma de los retornos proyectados para cada uno de sus componentes ponderados por la proporción de la cartera invertido en cada uno de ellos:

$$\hat{R}_{\text{portfolio}} = \sum_{i=1}^n \hat{R}_i \lambda_i$$

$$\text{Tal que: } \sum_{i=1}^n \lambda_i = 1$$

Donde \hat{R}_i representa el retorno proyectado por la RNA para el activo “i” y $\hat{R}_{portfolio}$ Corresponde al retorno proyectado para la cartera de cumplirse lo proyectado por la red.

Segundo, como indicador del riesgo se usará una medida de cuánto se equivocó al tratar de hacer las proyecciones para el signo de la variación del precio de cada acción. Esta medida será:

$$Riesgo = 1 - PPS$$

Donde “PPS” corresponde al Porcentaje de Predicción de Signo, que es una medida de en que porcentaje de las proyecciones hechas por la red esta acertó al signo del cambio del precio de la acción. Por lo que 1-PPS corresponderá al porcentaje en que se equivocó.

Nuevamente el riesgo será la suma ponderada de los riesgos de cada acción por el porcentaje de la cartera invertida en ellas. Si al riesgo del portfolio lo llamamos $\gamma_{portfolio}$ queda:

$$\gamma_{portfolio} = \sum_{i=1}^n (1 - PPS_i) \lambda_i$$

$$Tal \quad que : \sum_{i=1}^n \lambda_i = 1$$

Donde PPS_i corresponde al Porcentaje de Predicción de Signo de la red para la acción “i”.

Nuevamente se maximiza el retorno proyectado de la cartera, sujeto a un nivel dado de riesgo que esté dispuesto a soportar el inversionista, esa vez la expresión del problema de maximización queda:

$$\text{Máx } \widehat{R}_{portfolio} \quad \text{Sujeto a : } \gamma_{portfolio} = \gamma^*$$

Donde γ^* corresponde al nivel de riesgo (nivel de error de la red) que está dispuesto a soportar el inversionista para su cartera.

Además debo mencionar que a la hora de conformar portfolios por esta metodología establecimos 2 cotas para el riesgo, una denominada “riesgo al alza” y otra “riesgo a la baja”. El riesgo “al alza” es el máximo riesgo que estaba dispuesto a soportar si el mercado venía al alza y el riesgo “a la baja” era el máximo para el caso contrario. De esta forma se le permitía al conformador de carteras soportar un mayor riesgo si el mercado venía al alza y uno menor si venía a la baja.

II.1.5 RESULTADOS COMPARABLES

Para poder comparar la rentabilidad de ambas metodologías y ver cuál obtiene la mayor, se debe lograr que ambos tengan los mismos niveles de riesgo. Porque si tuviesen distintos riesgos no se podrían comparar y no se podría determinar algún tipo de dominancia de una técnica sobre otra. Para poder lograr esto se calculó el beta de cada portfolio conformado por RNA y después se forzó que las carteras formadas por el otro método tuviesen el mismo beta para las mismas semanas. Al hacer esto no estaríamos “comparando peras con manzanas”.

II.1.6 VALIDACIÓN Y CAPACIDAD PREDICTIVA DE LOS MODELOS

Para probar la capacidad predictiva de la red Ward con metodología Rolling para cada acción se usó el Test de Precisión Direccional de Pesaran & Timmermann (Llamado Test DA, por Directional Accuracy). Esta prueba es especial para los modelos donde lo que importa es la capacidad de predecir el signo de la variación de la incógnita.

Lo que esta prueba hace es testear si la proyección de la red y las observaciones reales están independientemente distribuidas. Para eso compara el PPS (llamado acá SR por Success Ratio), logrado por la proyección, con el PPS que habría logrado si las proyecciones y los datos observados fuesen independientes (Este PPS se llama SRI). De esta forma la fórmula para el test DA es:

$$DA = \frac{(SR - SRI)}{\sqrt{Var(SR) - Var(SRI)}}$$

DA sigue una distribución normal, por lo que para probar la hipótesis nula de que las observaciones y las proyecciones están independientemente distribuidas el resultado de esta ecuación se compara con un t crítico, que dependerá del nivel de confianza que se le quiera exigir a la prueba.

Además el resto de las variables se define como:

$$SR = PPS = \frac{\text{Cantidad de signos acertados}}{N^{\circ} \text{ de Predicciones}} = \frac{\sum_{i=1}^n I_i(y^* \hat{y} > 0)}{m}$$

Donde I_i corresponde a una función dicotómica que toma el valor 1 cuando la red acierta al signo del cambio de la variable y 0 cuando se equivoca en la predicción. “m” es la cantidad total de predicciones que se hicieron.

Además, el Ratio de Éxito cuando las observaciones y las proyecciones están independientemente distribuidas (SRI) se defina como:

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

Donde:

$$P = \frac{\text{signos} > 0 \text{ observados}}{\text{datos} \text{ totales}} = \frac{\sum_{i=1}^n I_i(Y > 0)}{m}$$

$$\hat{P} = \frac{\text{signos} > 0 \text{ proyectados}}{\text{datos} \text{ totales}} = \frac{\sum_{i=1}^n I_i(\hat{Y} > 0)}{m}$$

Que podría ser considerado como la “seudo – probabilidad” de que una de las proyecciones (o observaciones en el caso de P) sea positiva.

Por otro lado tenemos que las varianzas de SRI y SR son:

$$Var(SRI) = \frac{\left[m(2\hat{P} - 1)^2 P(1 - P) + m(2P - 1)^2 \hat{P}(1 - \hat{P}) + 4P\hat{P}(1 - P)(1 - \hat{P}) \right]}{m^2}$$

$$Var(SR) = \frac{SRI(1 - SRI)}{m}$$

Así el test DA de Pesaran & Timmermann, asumiendo que el test (o los resultados de este) siguen una distribución normal, prueba la hipótesis nula de que

los Y (observaciones) están independientemente distribuidos de los \hat{Y} (proyecciones). Por lo que si se rechaza la hipótesis nula se puede decir que existe evidencia de que se cuenta con capacidad para predecir la evolución de la variable observada. Así, para este estudio, le exigiremos un 95% de confianza a la prueba. Con esto, para rechazar la hipótesis nula, el resultado del test DA debe ser mayor (en valor absoluto) a 1.96.

II.1.7 DATOS

Para este trabajo se tomaron precios de cierre ajustados semanales para 29 acciones del índice *Dow Jones Industrial* y valores del *Dow Jones Industrial Average* para todos los días viernes comprendidos entre el 4 de Febrero de 1994 y el 3 de Septiembre de 2004. Estos datos fueron descargados del sitio “Yahoo! Finance”.

Las acciones usadas fueron: Alcoa Inc. (AA), American International Group Inc. (AIG), American Express Co. (AXP), Boeing Co. (BA), Citigroup (C), Caterpillar Inc. (CAT), Du Pont Co. (DD), Walt Disney Co. (DIS), General Electric Co. (GE), General Motors (GM), Home Depot Inc. (HD), Hewlett-Packard (HPQ), International Business Machines (IBM), Intel Corp (INTC), Johnson & Johnson (JNJ), JP Morgan Chase (JPM), Coca-Cola Co. (KO), McDonalds Corp (MCD), 3M Company (MMM), Altria Group (MO), Merck & Co. (MRK), Microsoft Corp. (MSFT), Pfizer Inc. (PFE), Procter & Gamble (PG), SBC Communications Inc. (SBC), United Tech Corp. (UTX), Verizon Communications (VZ), Wal-Mart Stores (WMT) y Exxon Mobil Corp. (XOM).

Estos datos fueron usados para entrenar, probar la calidad de ajuste y predecir los valores de cada semana requerida. Se requerían 10 años de datos (520 datos) para poder ir re – entrenando la red para cada una de las 190 semanas que se predijeron.

Para los días viernes que no existía información, simplemente se omitía el dato, reemplazándolo por un asterisco, para tratar de modificar lo menos posible los datos reales obtenidos.

Para cada predicción se usaron 300 datos en el conjunto de entrenamiento y 30 datos en el conjunto de producción⁶. No usamos un conjunto de prueba ya que podíamos calcular el ECM de la red cuando trataba de predecir un dato nuevo, ya que estos datos son pasados por lo que ya teníamos los valores reales de las variaciones de los precios de las acciones. Así, se puede ver que usamos 520 datos en total, contando todas las predicciones para las 190 semanas⁷

⁶ Conjunto de producción se denomina el conjunto de datos que recibe la red neuronal para la predicción del retorno de la semana de interés.

⁷ 520 datos totales – 300 datos para entrenamiento – 30 datos para producir = 190 semanas predecibles.

Parte III

III.1 RESULTADOS

III.1.1 CAPACIDAD PREDICTIVA

En la **tabla 1** se puede ver el resumen de las estadísticas para la capacidad predictiva del modelo para cada acción.

Acción	PPS	
	Extram.	Test DA*
1 - AA	71,233%	6,4060
2 - AIG	60,731%	3,2162
3 - AXP	59,361%	2,7723
4 - BA	64,384%	4,2238
5 - C	70,320%	5,9937
6 - CAT	70,776%	6,2052
7 - DD	68,493%	5,4578
8 - DIS	61,644%	3,4801
9 - GE	66,210%	4,9090
10 - GM	62,100%	3,4739
11 - HD	66,667%	5,0547
12 - HPQ	68,037%	5,5635
13 - IBM	69,406%	5,8573
14 - INTC	69,863%	5,9474
15 - JNJ	64,384%	4,3804
16 - JPM	63,927%	4,2538
17 - KO	73,059%	6,6762
18 - MCD	70,320%	6,0233
19 - MMM	69,863%	5,7186
20 - MO	73,973%	6,9082
21 - MRK	68,037%	5,3358
22 - MSFT	74,886%	7,2679
23 - PFE	73,516%	7,0589
24 - PG	75,799%	7,4723
25 - SBC	69,406%	5,7244
26 - UTX	70,320%	5,7914
27 - VZ	75,799%	7,6300
28 - WNT	72,603%	6,8110
29 - XOM	67,580%	5,1334
Promedio	68,714%	5,5430

* t-crítico=1,96

Tabla 1: Resultados PPS y Test DA

En la tabla se puede ver que el menor PPS fue de 59,361% para AXP (American Express Co.) y el mejor fue de 75,799% para PG (Procter & Gamble). A su vez, el modelo logró en promedio un PPS de 68,714% entre todas las acciones. O sea que en promedio acertó a más de la mitad de los signos de la variación de las acciones.

Por otro lado, los resultados del test DA muestran que para todas las acciones se rechaza la hipótesis nula, por lo que se puede decir que tenemos capacidad predictiva para todos los activos de la muestra.

III.1.2 PORTFOLIOS

A continuación se presentan los resultados de la conformación de portfolios tanto por la metodología tradicional (betas) como por la propuesta en esta tesis (RNA) en comparación con la rentabilidad de un portfolio Buy & Hold en el que se invierte igual proporción de los recursos en cada una de las acciones utilizadas para este trabajo (o sea un 3,45% en cada una).

En la **tabla 2** se muestran los resultados de la rentabilidad acumulada para portfolios donde se obliga a invertir el 100% de la riqueza del inversionista y para los que se permite realizar ventas cortas, pero no se permite apalancamiento. También se muestra en la última columna la cantidad de semanas en que cada método le ganó al otro (Betas o RNA). Acá se presentan resultados para portfolios que permiten concentraciones máximas de 10% y 15% por activo para las metodologías de Betas y RNA.

Método	Concentración Máx.	Promedio	Desvest.	Máximo	Mínimo	Semanas Ganadoras
Betas	10,00%	80,656%	3,425%	95,872%	68,512%	68
RNA	10,00%	81,484%	4,476%	94,720%	44,255%	124
Betas	15,00%	91,015%	3,984%	109,082%	79,386%	68
RNA	15,00%	92,684%	5,720%	147,050%	72,734%	124
Buy & Hold	3,45%	-9,537%	8,282%	6,264%	-29,215%	0

Tabla 2: Resultados para la rentabilidad acumulada de carteras con venta corta

Se puede ver que para ambas concentraciones máximas los portfolios conformados por RNA logran una mayor rentabilidad acumulada promedio que los portfolios de Betas, aún cuando esta ventaja es de alrededor de un 1% en ambos casos, estas son extremadamente superiores al promedio que alcanzó el portfollio “ingenuo” o “Buy & Hold”, rentabilidad que no solo es pequeña en valor absoluto, si no que además es negativa para el período analizado, esto se explica debido al mal momento que pasó el Dow Jones durante la ventana de tiempo estudiada⁸. Así, como el portfollio ingenuo se conforma con 29 de las 30 acciones del Dow, es fácil ver que éste seguirá la tendencia del índice de mercado usado.

En la **tabla 2** también se puede ver que los portfolios conformados por RNA superaron en términos de rentabilidad acumulada a los portfolios por Betas en 124 ocasiones, casi el doble de veces en que los portfolios de betas superaron a los de RNA. En esta misma columna se puede ver que el portfollio de Buy & Hold nunca superó a ninguno de los otros dos métodos en términos de rentabilidad acumulada.

También se aprecia que la desviación estándar siempre es mayor con la metodología de RNA que con Betas. Pero la desviación estándar del portfollio Buy & Hold es considerablemente mayor que la de los otros dos métodos.

Ahora en la **tabla 3** se pueden ver los mismos resultados que en la tabla anterior, sólo que esta vez no se permite la venta corta.

Método	Concentración Máx	Promedio	Desvest	Máximo	Mínimo	Semanas Ganadoras
Betas	10,00%	-3,013%	3,516%	7,474%	-27,366%	109
RNA	10,00%	-6,781%	9,664%	8,302%	-44,225%	83
Betas	15,00%	-1,544%	3,339%	9,817%	-23,902%	98
RNA	15,00%	-3,609%	8,414%	9,957%	-44,464%	94
Buy & Hold	3,45%	-9,537%	8,282%	6,264%	-29,215%	48

Tabla 3: Resultados para la rentabilidad acumulada de carteras sin venta corta

⁸ Ver anexo 1: “Evolución del Dow Jones diciembre 2000 a diciembre 2004”.

Al prohibir la venta corta los resultados son totalmente distintos, aquí todos los portfolios obtuvieron rentabilidades negativas en promedio. Además, los portfolios por Betas obtienen mayor rentabilidad acumulada promedio para ambas concentraciones que los portfolios por RNA. Por lo menos ambas metodologías siguen obteniendo mejores resultados promedio que el portfolio ingenuo (que como ya dijimos es virtualmente el de mercado), ya que ninguno de ellos pierde tanto dinero como el mercado en el período en cuestión.

Ahora se aprecia que los portfolios de Betas superaron en rentabilidad acumulada a los de RNA en mayor cantidad de ocasiones para ambas concentraciones, pero el margen es mucho menor que el logrado por los portfolios de RNA cuando se permitía la venta corta. Además se puede ver que esta vez si existieron semanas en que el portfolio ingenuo superó a los otros 2 métodos, de hecho lo logró en un no despreciable 25% de las semanas estudiadas.

Las desviaciones estándar siguen siendo mayores para los portfolios de RNA que para los de Betas, pero ahora existe un caso en que la desviación estándar de los portfolios por RNA superó a la del portfolio de Buy & Hold, cosa que no sucedió cuando se permitía hacer venta corta.

Se puede apreciar que la diferencia entre la situación con y sin venta corta es abismante en términos de la rentabilidad acumulada lograda por ambos métodos, era de esperarse que este indicador fuese mayor en el caso de estar permitida la venta corta. Lo que no es tan intuitivo es el hecho de que con venta corta el portfolio por RNA logre mejor rendimiento que el de Betas y que sin venta corta la situación sea a la inversa.

Parte IV

IV.1 CONCLUSIONES

La red Ward de 3 capas usada en este estudio posee una considerable capacidad predictiva para el signo que seguirá la variación del precio de un activo, esto la convierte en una herramienta muy potente para la generación de retornos proyectados más creíbles y por ende para conformar carteras de inversión. Así, las redes neuronales se pueden convertir en una opción a las metodologías tradicionales de conformación de carteras, manteniendo la misma idea básica de maximizar el retorno sujeto a un riesgo dado, pero usando medidas alternativas a las conocidas, tanto para el riesgo como para el retorno esperado (o proyectado en el caso de RNA).

La red neuronal usada en este estudio es especialmente efectiva cuando se permite la realización de ventas cortas, logrando una rentabilidad considerable al cabo de 4 años (aproximadamente). Sin embargo, cuando no se permite hacer venta corta, la metodología de conformación de carteras por betas logra un mejor desempeño en promedio. Esto no es intuitivo, pero se podría deber a las concentraciones máximas con las que se obligó a trabajar a las metodologías, quizás si no estuviesen tan limitadas las concentraciones, se lograrían rentabilidades positivas con ambos métodos. Además la estrategia con RNA siempre generó mayores costos totales de transacción, sobre todo para los portfolios sin venta corta, lo que llevó a que la estrategia por RNA tuviera menor rentabilidad acumulada promedio. Pero en el caso de los portfolios con venta corta la Red neuronal logró superar a los Betas en casi el doble de ocasiones, siendo que en los portfolios sin venta corta ambas metodologías se superaron la una a la otra en casi igual cantidad de veces. Aún cuando la de betas venció a la de RNA en mayor número de semanas, la supremacía en este caso no es tan pronunciada como en el anterior. Sin embargo ambas metodologías logran “ganarle al mercado” al obtener mayores rentabilidades que este, aún cuando no logran evitar obtener rentabilidades negativas sin el uso de ventas cortas ni bajo las

restricciones impuestas para la concentración máxima de cada acción permitida en las carteras.

De lo anterior se puede decir que cuando se permite la venta corta de acciones, la RNA ejerce dominancia estocástica sobre la estrategia de Betas, pero si no se permite venta corta, es a la inversa. Esto se puede deber exclusivamente a las características económicas del período analizado, junto con las restricciones impuestas sobre la concentración máxima en cada activo de la cartera.

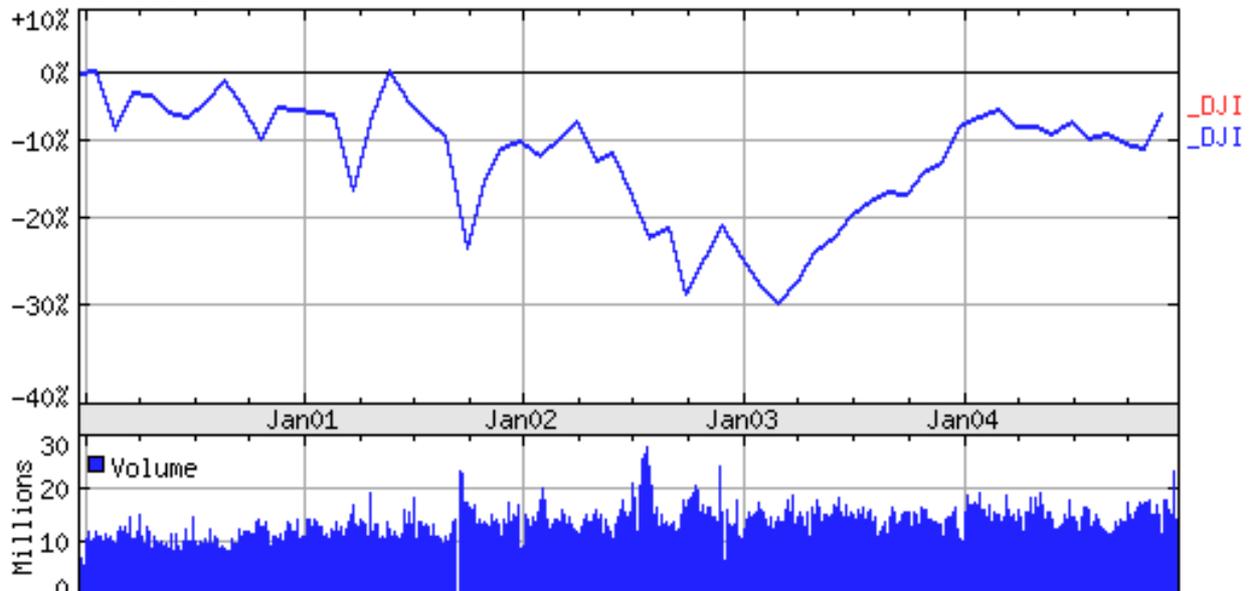
También se puede concluir que al aumentar la proporción máxima permitida para invertir en cada acción la rentabilidad de ambas metodologías comparadas aumentó, lo que era de esperarse. Sin embargo la desviación estándar aumentó considerablemente al pasar de portfolios con venta corta a los sin venta corta, cosa que no ocurrió para las carteras conformadas por betas. Esto indica que la estrategia por RNA tiene mayor volatilidad en sus retornos si no se permite la venta corta de acciones.

Las redes neuronales pueden constituir una poderosa herramienta para la proyección de retornos para la conformación de carteras, pero al igual que con las otras metodologías para conformar portfolios, hay que tener en cuenta que el factor humano siempre seguirá siendo fundamental, ya que las finanzas y la economía y sobre todo los precios accionarios tienen tras de ellos a seres humanos, por lo que siempre se verán afectados por muchos otros factores aparte de la evolución del mercado, cosas inesperadas como un atentado terrorista, el despido o la renuncia de una persona clave en una organización, el anuncio de pago o no pago de dividendos, el aumento del precio de los commodities o incluso que se le descubra un amorío al Gerente General de una empresa o que este se divorcie y tantos otros sucesos que no se pueden anticipar por medios estadísticos ni analíticos, en estos casos solo un humano es capaz (hasta ahora) de incorporar esta información y predecir como impactará el precio de un activo. Por lo que seguirá siendo importante la supervisión por parte de un analista humano.

Anexo 1:

***“Evolución del Dow Jones diciembre 2000 a
diciembre 2004”***

DJ INDU AVERAGE (Dow Jones & Co
as of 20-Dec-2004



Copyright 2004 Yahoo! Inc.

<http://finance.yahoo.com/>

Acá se puede apreciar la evolución del Dow Jones Industrial Average para el período cubierto en este trabajo. Del gráfico se desprende que este tuvo una rentabilidad acumulada negativa para la ventana de tiempo que nos interesa y de ahí que un portfollio tipo Buy & Hold que se conforme en diciembre del 2000 con un 3.45% de cada acción del Dow y se mantenga hasta septiembre del 2004 habría tenido una rentabilidad acumulada negativa.

REFERENCIAS

Markowitz, H.M. *“Portfolio selection”*. Journal of finance Vol. 7. 1952.

Markowitz, H.M. *“Portfolio selection: efficient diversification of investments”*. John Wiley and Sons Inc. 1959.

Fama, Eugene F. *“Efficient capital markets: a review of theory and empirical work.”* Journal of finance. Mayo 1970.

Fama, Eugene F. *“Efficient capital markets: II”*. Journal of finance. December 1991.

Campbell, J.; Lo, A.; MacKinlay, A. *“The econometrics of financial markets”*. Princeton University Press. 1997.

Gavridis, M. *“Modelling with high frequency data: a growing interest for financial economists and fund managers”*. Nonlinear Modelling of High Frequency Financial Time Series. Dunis, C. Zhou, B. Editores. 1998.

Lo, A. y MacKinlay, A. C. *“Stock market prices do not follow random walk: Evidence from simple specification test.”* Review of financial studies. 1988.

Ferson, W y Harvey, C. *The variation of economic risk premiums”*. Journal of Political Economy 99. 1991

Leung, M. T.; Daouk, H y Cheng, A. *“Forecasting stock indices: a comparison of classification and level estimation models”*. International Journal of Forecasting 16. 2000.

Martín del Brío, B. y Sanz Molina, A. *“Redes neuronales y sistemas borrosos”*. Editorial RA-MA, Madrid. 1997.

Reyes, Daniel y Tapia, Victor. *“Evaluación de los distintos modelos de redes neuronales en la predicción de valores financieros”*. Facultad de Ciencias Económicas y Administrativas, Universidad de Chile. Santiago, Chile. 2000.

Parisi, Antonino. Parisi, Franco y Guerrero, José Luis. *“Redes neuronales en la predicción de Índices bursátiles internacionales: Un análisis de la estabilidad de los pesos”*. El trimestre económico. México. 2003.

Pesaran, M. H. y Timmermann, A. *“A simple non parametric test of predictive performance”*. Journal of Business and Economic Statistics. 1992.