MODELO DE CLASIFICACIÓN PARAMÉTRICA Y NO PARAMÉTRICA DE LOS FENÓMENOS DE RIESGO BANCARIO EN VENEZUELA

Alexis A. Melo T., Gerardo Colmenares

Resumen

Este proyecto en particular trata con un modelo híbrido en el que se aprovecha las bondades de la técnica multivariante lineal, conocida como Análisis Discriminante (AD) y el aprendizaje cualitativo no supervisado mediante la función base radial, bajo la topología de redes neuronales para la gestión de riesgo bancario. Este modelo se configura en dos etapas: la primera, para determinar los patrones de salida mediante AD y la segunda, para determinar un modelo de pronóstico y a su vez, de reconocimiento de las agrupaciones de bancos que se originan en la capa oculta, luego de acceder a los patrones de entrada que son las razones financieras transformada a una escala, previamente. El desempeño de este modelo está siendo evaluado mediante el uso del conjunto construido de datos para tales fines. El objetivo fundamental de este proyecto, es conjugar ambas técnicas para obtener un mecanismo automático de clasificación de riesgo en la banca y la eventual capacidad de generalizar y poder pronosticar el comportamiento de un banco en particular, dado que se conozca el patrón de entrada.

Palabras claves: Entrenamiento supervisado, análisis discriminante, redes neuronales, función base radial, clasificación de riesgo bancario.

Introducción.

El razonamiento cualitativo es una de las habilidades del aprendizaje en los modelos de redes neuronales y formaliza la capacidad humana de percibir, analizar, entender y modelar problemas reales. En el caso particular de intentar clasificar el riesgo, el proceso de aprendizaje cualitativo puede ser crucial debido a que la descripción del valor de los patrones de entrada puede aparecer en diferentes escalas. Pueden influir variables tanto numéricas como no numéricas, tal como podría ocurrir en el aprendizaje humano.

Eventualmente esta técnica ha sido empleada para pronóstico de riesgo de crédito cuando son clasificados los bonos de una compañía particular. (Rovira et al.). Indistintamente se han considerado como variables influyentes tanto variables categóricas como variables cuantitativas ajustadas a una escala.

Un caso similar ocurre con la investigación que se desarrolla en este proyecto. El empleo de las razones financieras como patrones de entrada y los patrones de salida por cada valor observado como patrón de entrada estimado mediante la clasificación obtenida por medio de un procedimiento de análisis discriminante, determina el dominio del problema.

Las redes neuronales basadas en funciones de base radial (RBF) contribuyen a emular este tipo de condición de los datos. El desempeño de clasificación interna en grupos que podría realizar el modelo RBF en su capa oculta de neuronas, estaría inicialmente influenciado por la función empleada capaz de atraer patrones a un grupo determinado y consecuentemente la presentación de los datos descritos en los patrones de entrada y salida. Algunos estudios enfatizan que mientras sea más cualitativa la información, mejor desempeño tendría el modelo. (Rovira et al., 2002). Por otro lado, (Atiya, 2001) señala en su revisión profusa que en general las redes neuronales han mostrado mejor desempeño en el pronóstico en el dominio de problemas financieros. Sin embargo, el enfoque presentado en este proyecto pareciera ser un enfoque emergente al proponerse la capacidad del modelo híbrido de: a) pronosticar una situación en la banca comercial para un banco determinado, b) agrupar la banca comercial de acuerdo a las similitudes en los patrones de entrada, c) combinar ambas técnicas, AD y RBF, en el proceso de preprocesamiento de datos, proceso de aprendizaje de la red y capacidad de pronóstico.

En las secciones que siguen, se describirán brevemente las técnicas que se emplearán y finalmente el modelo que se esta siguiendo para este proyecto en particular.

Análisis Discriminante aplicado al riesgo bancario

Teniéndose un amplio conjunto de datos p-dimensionales el problema del Análisis Discriminante se reduce a la habilidad que se tiene, mediante un modelo lineal, de poder clasificar este conjunto de datos de entrada con sus variables definidas, en función de una regla de discriminación o de clasificación predeterminada.

Reforzando la idea anterior (Sharma, 1996) resumió la aplicabilidad de este método en los siguientes tres objetivos:

- Identificar un conjunto de variables que *mejor* discrimine o separe entre los grupos.
- Identificar un nuevo eje Z, tal que, la nueva variable Z, dada por la proyección de las variables originales sobre este eje, provea la máxima separación o discriminación entre 2 grupos, vale decir, un hiperplano que permita la máxima separación entre grupos para el caso multidimensional.
- Clasificar futuras observaciones dentro de uno de los grupos. Es decir su capacidad de pronóstico.

El Análisis Discriminante entonces, por lo anterior, entra entre las técnicas definidas como Clasificación Supervisada, pues a partir de una muestra de observaciones previamente bien clasificados se permite la clasificación de nuevos datos adicionales.

Este método es de amplia aplicación en diversos campos, y para el área financiera tiene especial aplicabilidad en el análisis del credit-scoring y del estudio del problema de la turbulencia financiera en el sistema bancario. (Altman et al, 1994); (Atiya, 2001); (Foster et al, 1999).

El Análisis Discriminante Clásico se conoce como el Método de Discriminación de Fisher en honor a quien inicialmente lo desarrollo partiendo de una relación lineal de discriminación. Esta relación de discriminación lineal viene dada por la siguiente expresión $\mathbf{Z}=\mathbf{X}'\boldsymbol{\gamma}$, donde $\mathbf{X}$ representa el vector de variables observadas, $\boldsymbol{\gamma}$ representa los pesos de la función discriminante y $\mathbf{Z}$ es la función discriminante. De lo anterior se describe que $\mathbf{Z}$ resultará como una variable dependiente indicadora y no numérica.

La variable indicadora puede tomar dos o más valores, no obstante, el caso más común es el que emplea dos valores: cero y uno, en este caso se conoce como variable dicotómica.

Por ejemplo, la variable $\mathbf{Z}$ podría ser una variable indicadora, que denota la ocurrencia o no de un evento.

En este contexto, el Análisis Discriminante se emplea para determinar cuál o cuáles variables contribuyen a discriminar entre dos o más grupos que se observan en la práctica.

En resumen, la idea básica que subyace en el Análisis Discriminante es determinar si unos grupos difieren en función de la media de una variable, y emplear luego esa variable para predecir la pertenencia de una nueva observación a determinado grupo.

El problema del Análisis Discriminante, visto desde el punto de vista del análisis de variancia, consiste en responder a la pregunta de si dos o más grupos son significativamente diferentes uno de otro respecto a la media de una variable en particular.

Debe tenerse presente que si la media de una variable es significativamente diferente en varios grupos, puede decirse que esta variable discrimina entre grupos.

Al igual que en el caso de conglomerados, análisis factorial y de correlación canónica, el Análisis Discriminante realiza diferentes desgloses de las variancias de un conjunto de datos para someterlos a una serie de pruebas estadísticas y determinar el grado de asociación entre esas variancias y, por tanto, entre las variables. De esta forma, la mejor discriminación se tiene al maximizar la relación

$$\frac{\mathrm{Var}(\mathbf{Z})\_\mathrm{entre}\_\mathrm{grupos}}{\mathrm{Var}(\mathbf{Z})\_\mathrm{dentro}\_\mathrm{del}\_\mathrm{grupo}}$$

En el caso de una única variable explicativa, la prueba final de significancia de si esta variable discrimina o no entre grupos es una prueba F, que es básicamente una razón de las variancias entre grupos sobre el promedio de la variancia dentro de los grupos. Si la variancia entre grupos es significativamente mayor, deberá haber diferencias significativas entre las medias.

Para el caso de más de una variable, se busca determinar cuál o cuáles de ellas contribuyen a la discriminación entre grupos. En este caso, se tiene una matriz de variancias y covarianzas. Puede compararse las matrices con una prueba F multivariable, para determinar si hay o no diferencias significativas en las medias entre grupos.

En el caso particular de función discriminante para dos grupos, tal función puede verse como una caso de análisis de regresión lineal múltiple. Si se codifican los dos grupos como 1 y 2, y se emplea tal variable como dependiente en un análisis de regresión lineal múltiple, pueden obtenerse resultados similares a los que se obtendrían de un análisis discriminante. En general, en el caso de dos grupos se ajusta una ecuación lineal del tipo

$$\mathrm{GRUPO} = a + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_m x_m,$$

donde *a* es una constante y b_1 a b_m son coeficientes de regresión. La interpretación de estos resultados es similar a la de un modelo de regresión múltiple. Los más significativos son lo que contribuyen más a la predicción de pertenencia a un grupo.

Para efectuar el análisis, es posible emplear diferentes procedimientos, como por ejemplo:

Análisis discriminante "Stepwise": probablemente la forma más común de aplicación es incluir muchas medidas en el estudio, para determinar las que discriminan entre grupos. Visto de otra forma, se desea construir un modelo de cómo se puede lograr predecir de la mejor forma a cuál grupo pertenece una observación o caso particular.

Análisis discriminante "Stepwise" hacia adelante: la idea es construir un modelo paso a paso, revisando todas las variables y evaluando cuál puede contribuir más a la discriminación entre grupos. Esta variable podrá ser incluida en el modelo.

Análisis discriminante "Stepwise" hacia atrás: es posible incluir primero todas las posibles variables en el modelo, y luego en cada paso, eliminar la variable que contribuye menos a la predicción de la pertenencia a un grupo. Como resultado de un modelo de Función Discriminante exitoso, deben mantenerse en él las variables significativas para discriminar.

Cuando es posible identificar más de dos grupos, puede estimarse más de una función discriminante similares a la presentada anteriormente. Por ejemplo, cuando se tienen tres grupos, puede estimarse: a) una función para discriminar entre grupo 1 y grupos 2 y 3 combinados, y b) otra función para discriminar entre grupo 2 y grupo 3.

En la práctica, cuando se realiza un análisis discriminante entre varios grupos, no debe especificarse cómo combinar los grupos para formar las diferentes funciones. Los programas computacionales automáticamente las conforman de manera que la primera es la que ofrece la mayor discriminación como un todo entre grupos, la segunda provee una menor y así sucesivamente. Las funciones son independientes u ortogonales, esto es, su contribución a la discriminación entre grupos no se sobrepone.

Puede probarse el número de variables que agregan significancia a la discriminación entre grupos. Solo aquellas que sean estadísticamente significativas deben ser usadas para interpretar, las no significativas deben ignorarse.

En resumen, cuando se interpretan funciones discriminantes múltiples, que surgen del análisis con más de dos grupos y más de una variable, se puede probar primero la significancia estadística de las diferentes funciones, y considerar solo las significativas para las siguientes pruebas. Luego, se observan los coeficientes *b* estandarizados (expresión anterior) para cada variable y para cada función significativa. Cuantos mayores sean, más alta es la contribución a la discriminación especificada por la respectiva función. Finalmente, pueden verse las medias para las funciones discriminantes significativas para determinar entre cuáles grupos discrimina la respectiva función.

Es importante tener presente unos supuestos implícitos a este tipo de análisis:

> *Distribución normal:* se asume que los datos para las variables representan una muestra proveniente de una distribución normal multivariable. No obstante, el no cumplimiento de este supuesto no es problema para el análisis.
>
> *Homogeneidad de varianzas y covarianzas*: se supone que las matrices de variancias y covarianzas son homogéneas entre grupos; de nuevo, si no se cumple tampoco se generan problemas.

Correlaciones entre medias y varianzas: el principal obstáculo para la validez de las pruebas de significancia se presenta cuando la media de las variables entre grupos están correlacionadas con las variancias. Si hay gran variabilidad en un grupo con alta media en algunas variables, entonces esas medias grandes no son confiables. Sin embargo, la prueba de significancia global está basada en variancias ponderadas, es decir en variancias promedio entre todos los grupos. La prueba de significancia de las medias relativamente grandes (con grandes variancias) estará basada en variancias ponderadas relativamente menores, resultando erróneamente en significancia estadística. Esto ocurre cuando un grupo tiene unos pocos valores extremos que afectan mucho la media y aumentan la variabilidad.

Variables no redundantes: se supone que las variables empleadas para discriminar entre grupos no son completamente redundantes, por ejemplo que una variable no sea la suma de otras dos que también están en el modelo.

En la figura siguiente se puede observar un esquema funcional del preprocesamiento de los datos financieros de la banca comercial basados en la aplicabilidad del análisis discriminante. Los valores obtenidos de la agrupación en dos grupos de la banca comercial de acuerdo a sus razones financieras, será empleado como patrones de salida en la construcción del modelo de red neuronal no supervisado usando función de base radial. El propósito de esta construcción de modelo es el lograr un modelo con un alto grado de aprendizaje que se pueda generalizar a todo el dominio del problema real bajo el mismo contexto de las razones financieras empleadas para el entrenamiento. El resultado de este buen desempeño es un mecanismo de pronóstico en el cual un nuevo ejemplo no visto por la red en el tiempo pueda ser separado en uno de los grupos con un alto grado de aceptación.

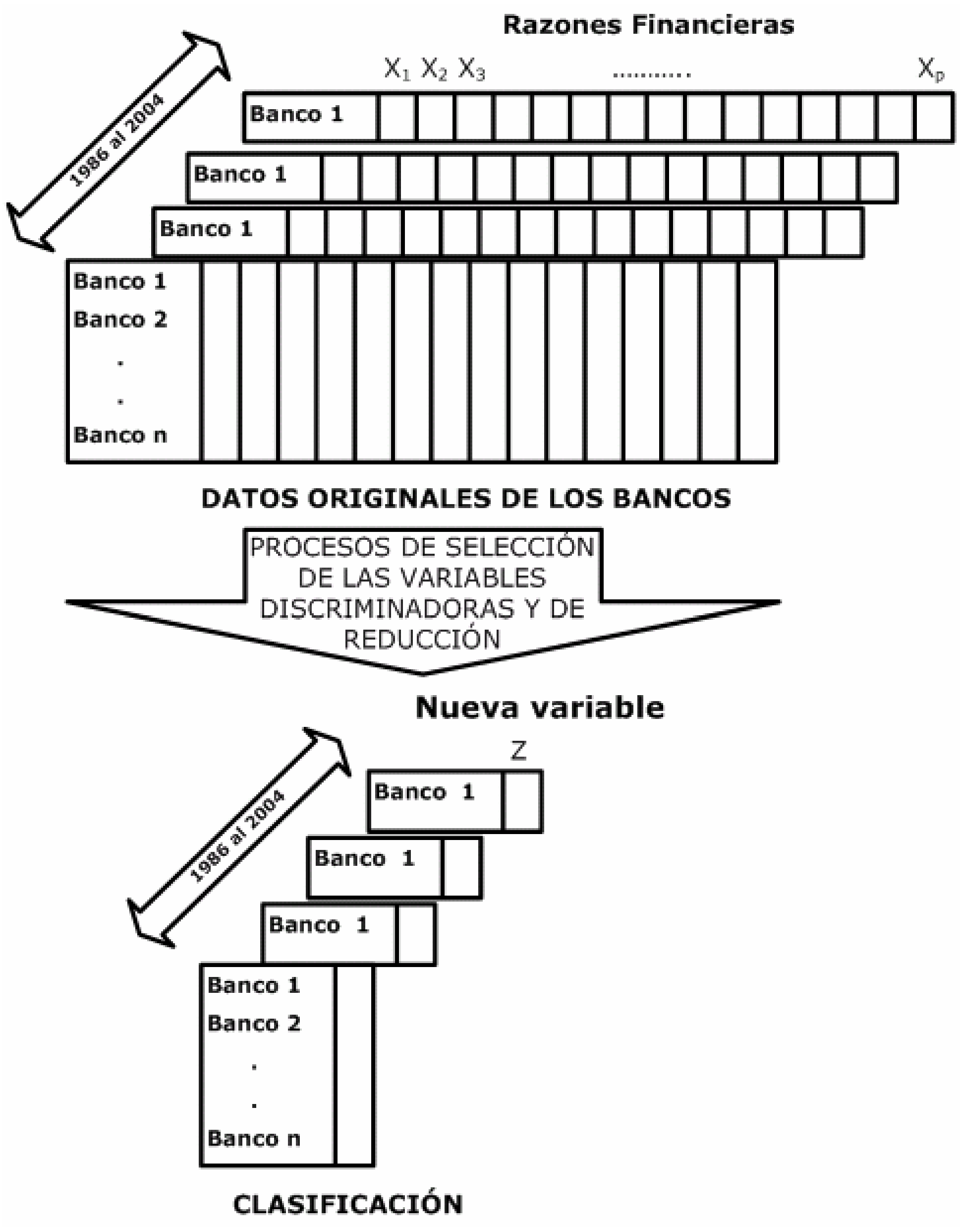

Figura 1. Discriminación de los bancos en dos grupos

Las Redes Neuronales Artificiales: su competencia como técnica de clasificación mediante función de base radial.

Una red RBF está conformada de tres capas: a) la capa de entrada que sirve para los ejemplos o patrones de entrenamiento y prueba, b) la capa oculta completamente interconectada entre todos sus nodos con la capa de entrada y activada a través de la función radial (gaussiana) y, c) la capa de salida, también completamente interconectada a la capa oculta y activada a través de una función lineal continua. Así, una red RBF tiene como objetivo el de ejecutar una *correspondencia no lineal* entre los patrones de entrenamiento que definen el espacio de entrada al espacio oculto definido por la capa oculta y una *correspondencia lineal* desde este espacio al espacio de salida. Es decir definir a la salida una superficie que describa las entradas.

El entrenamiento no supervisado, a diferencia de la red usando backpropagation (RN-RP), es solamente hacia delante. De este modo, la salida *z* de una red RBF, en general, está influenciada por una transformación no lineal originada en la capa oculta a través de la función radial y una lineal en la capa de salida a través de la función lineal continua.

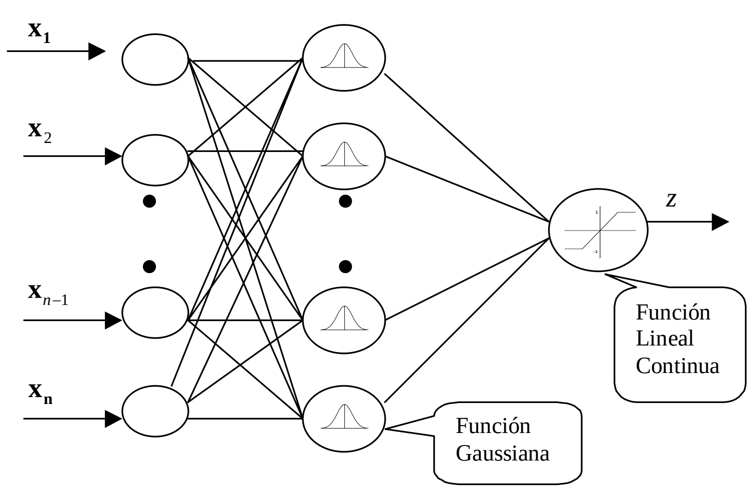

Figura 2. Arquitectura típica de una red neuronal con funciones de base radial

En la topología particular de una RBF se distinguen los nodos ocultos que contienen una función base radial o función gaussiana, la cual tiene como parámetros a un centro y un ancho. El centro es único para cada función radial involucrada en la capa oculta, y por otro lado, el ancho que identifica la amplitud de la campana de gauss originada por la función radial, es decir, la desviación estándar de la función radial. Algunos autores (Broomhead and Lowe, 1988) consideran a este ancho como un valor constante para cada una de las funciones radiales consideradas en la capa oculta y de este modo, contribuyendo a simplificar los pasos de construcción del modelo de entrenamiento de la red.

El primer cálculo efectuado en la capa oculta es hallar para cada nodo de la capa oculta la distancia radial (distancia euclidiana) *d* entre el vector de entrada *x* y el centro de gravedad *c* de ese mismo nodo, para cada una de las n observaciones. Es decir:

$$d = \|\mathbf{x} - \mathbf{c}\| = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \ldots\ldots + (x_n - c_n)^2}$$

Este valor *d* es calculado para cada nodo oculto y es un componente de la entrada para activar la función radial *G(•)* correspondiente a cada nodo. La función radial *G(•)* más comúnmente empleada es $exp(-r^2)$, donde ***r*** es el contenido evaluado en cada nodo de la capa oculta. En este caso particular, este contenido es la distancia euclidiana *d*. De ahí que la expresión anterior sería $exp(-d^2)$.

Una de las derivaciones del modelo RBF es emplear el ancho (desviación estándar) para activar la función *G(•)*. En este caso se estaría trabajando con algo como $exp(-d^2/a)$, donde *a* es el ancho para ese nodo oculto.

Entre la capa oculta y la capa de salida se derivan un conjunto de pesos **w** que se verían afectados de acuerdo al algoritmo de aprendizaje. En este caso particular. sería la combinación lineal entre los pesos y la resultante de cada función radial para determinar la salida *z*.

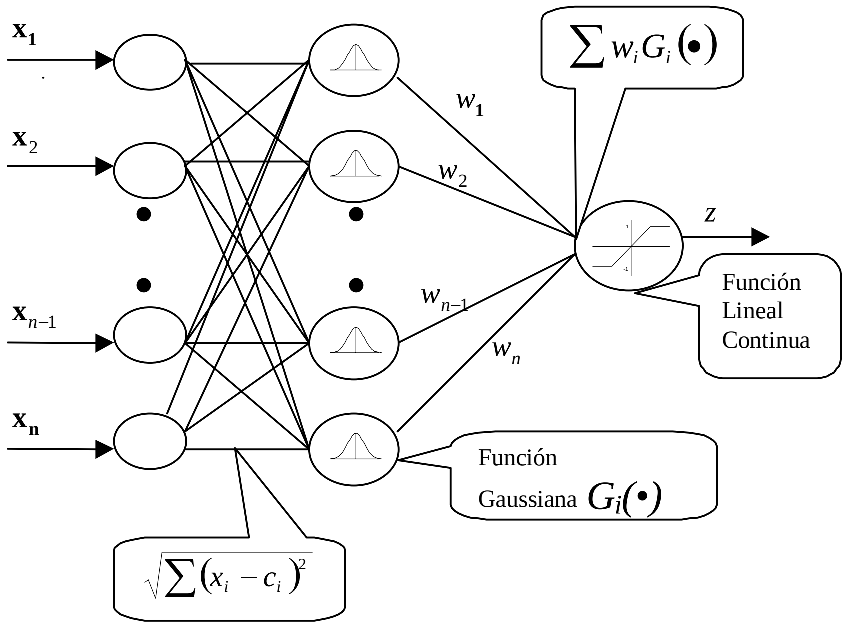

Figura 3. Componentes de una RBF: función de activación en ambas capas, pesos sinápticos

En definitiva, seria $z = \sum w_i G(\bullet)$, donde *G(•)* es la salida de la capa oculta y se corresponde con la función radial aplicada a la distancia euclidiana en cada una de las unidades ocultas.

Del resultado de este tipo de entrenamiento podemos observar que:

Los valores de entrada se recomiendan que previamente sean de algún modo transformados a una escala.

Debido a que esta superficie es desconocida, se acude un proceso de entrenamiento usando ejemplos representativos tanto para la entrada como para la salida. Es decir, muestras que incluyan ejemplos de todo el dominio del problema.

En la capa oculta, en la medida que los valores de entrada se parezcan más a un centro su distancia tenderá a *cero* y de este modo la función gaussiana se dispararía a las vecindades de *uno*. Por otro lado, en la medida que los valores de entrada no se parezcan a su centro la distancia será mayor y la función radial parecería tender a *cero*. Este proceso es una clasificación no lineal de las entradas.

En la capa de salida del modelo RBF, los valores obtenidos en las salidas de la capa oculta serían transformados por la función lineal que permite aproximar los valores *z* a los valores deseados, mediante la combinación lineal que se sucede en esta capa entre sus pesos y el resultado de aplicar la función radial. Es decir, $z = \sum w_i G(\bullet)$.

El tiempo de entrenamiento es substancialmente inferior al requerido por otros algoritmos. Es una pasada hacia adelante en la mayoría de los casos. La diferencia la establece si se incorpora en la salida del modelo de entrenamiento, una supervisión a través del control del error que se produce entre los valores calculados y los observados, conduciendo a una retropropagación del error.

De acuerdo a este concepto, la red RBF ha originado variantes de cálculo como producto fundamentalmente de las siguientes limitaciones que se han presentado en algunos casos:

de no conocer los centros (a veces el ancho) para cada función radial,

de situaciones de singularidad presentes en la implementación del algoritmo con problema de dimensionalidad;

de un gran volumen de entradas haciendo inmanejable la aplicación del algoritmo. Se presentan problemas de regularización (Haykin, 1995).

La función que puede describir sigue la siguiente forma:

$$F(\mathbf{x}) = \sum_{1}^{N} w_i G\left(\left\|\mathbf{x} - \mathbf{x_i}\right\|\right)$$

donde la función F(x) es una combinación lineal entre la función no lineal $G(\bullet)$ y los pesos. $G(\bullet)$ es de la forma $\exp(||\mathbf{x}_j-\mathbf{x}_i|)$ y la expresión anterior, en forma matricial sería **Gw = z.**

Cada elemento $g_{j,i} = g(||\mathbf{x}_j-\mathbf{x}_i||), j,i=1....N$

$\mathbf{z} = [\ z_1,\ z2\ z_3,\ ......,\ z_N]^T$

$\mathbf{w} = [w_1,\ w_2,\ w_3,\ .....,w_N]^T$

Provistos que todas las observaciones son distintas, entonces G se podría decir que es positiva definida y por lo tanto los pesos podrían ser calculados mediante la inversa de **G**. Es decir:

$$\mathbf{w} = \mathbf{G}^{-1}\mathbf{z}$$

Sin embargo, se puede correr el riesgo de que la inversa de la matriz de interpolación G está próxima a ser singular. En este caso se procedería mediante la teoría de la regularización para perturbar la matriz mediante $\mathbf{G} = \mathbf{G} + \lambda\mathbf{I}$. (Haykin, 1995)

De esta manera sería un aprendizaje directo, provocando cambio a los pesos que están ubicados entre la capa oculta y la capa de salida.

La aplicabilidad de ambas técnicas, análisis discriminante y redes neuronales de funciones de base radial, describen un modelo híbrido que incluye los datos construidos en una escala homogénea como patrones de entrada definidos por las razones financieras, y el resultado de utilizar el análisis discriminante para obtener los patrones de salida. La figura 4 describe funcionalmente la aplicabilidad de la RBF y en la figura 1 la construcción de los patrones de salida. Los patrones de entrada se construyen transformando las razones financieras en un nuevo conjunto de datos discretizados, tal que se pueda unificar las escalas de medida. Se aplicará el método de discretización supervisado CAIM (Class-Attribute Interdependence Maximization) (Ching et al., 1995); (Kurgan et al., 2001). CAIM es uno de los más reciente métodos de discretización dando muy buenos resultados en un tiempo de cómputo bastante reducido. Este método divide el rango de la variable, en este caso la razón financiera, en un número muy pequeño de intervalos que podría se encontrados automáticamente. Si la escala de intervalos se desea dejar fija, entonces se puede aplicar una variante del método (Campos et al., 2004).

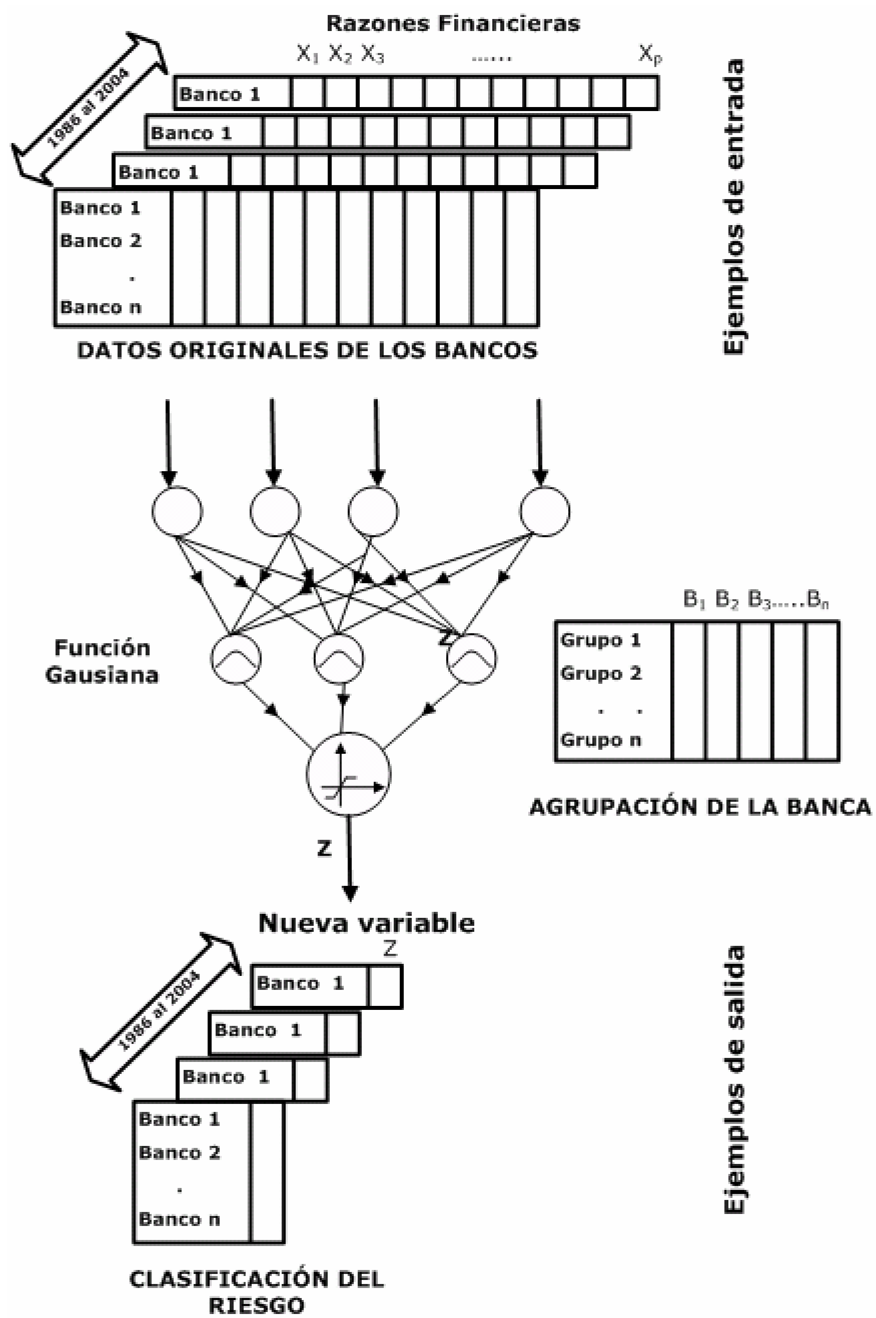

Figura 4. Pronóstico de riesgo clasificado en categorías

Referencias.

Agresti, A. (1990). Categorical Data Analysis. John Wiley and Sons Ltd., New York.

Amir F. Atiya, *Senior Member, IEEE.* (2001),Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 12, NO. 4, JULY 2001

Broomhead, D. S., and D. Lowe, (1988). Multivariable functional Interpolation and adaptive networks. *Complex Systems*, vol. 2, pp. 321–355.

Campos, R., Ruiz, F., Agell, N and Angulo, C. (2004). Financial Credit Risk Measurement Prediction Using Innovative Soft-computing Techniques. International Conference on Computational Finance And its Applications. Bologna, Italy, 2004.

Ching, J.Y., Wong, A.K.C. and Chan, K.C.C. (1995). Class-Dependent Discretization for Inductive Learning from Continuous and Mixed Mode Data, *IEEE Transactionson Pattern Analysis and Machine Intelligence*, 17 (7), pp. 641-651, 1995.

Colmenares G. and Pérez R. (1999). A Reliable Method to Reduce Observations and Variables when Building Neural Network Models. CAIP'99. San José. Costa Rica.

Hair, J., Anderson, R., Tatham, R. and Black, W. (1998). Multivariate Data Analysis. Prentice Hall, 5th Ed.

Haykin Simon. (1995). Neural Networks. A comprehensive foundation. Macmillan College Publishing Company, Inc.

Haykin Simon. (1998). Feedforward: Neural Networks: An Introduction. Chapter 1. Manuscrito

Kurgan, L. and Cios, K.J. (2001). Discretization Algorithm that Uses Class-Attribute Interdependence Maximization, Proc. of the 2001

International Conference on Artificial Intelligence (ICAI-2001): Las Vegas, pp.980-987, 2001

Muñoz S., Evelyn. (1998). La técnica de Análisis Discriminante: Una aplicación para el área bancaria. Banco Central de Costa Rica, DIE-NT-03-98

Sharma Subash. (1996). Applied Multivariate Techniques. John Wiley. USA

Xari Rovira, Núria Agell, Mónica Sánchez, Francesc Prats and Xavier Parra. (2002). An Approach to Qualitative Radial Basis Function Networks over Orders of Magnitude. This work was supported by the MCyT (Spanish Ministry of Science and Technology) MERITO project (TIC2002-04371-C02).

Wang Z. Xue (1999). Data Mining and Knowledge Discovery for Process Monitoring and Control. Springer-Verlag. London.