

Pronóstico de incumplimientos de pago mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito

Por:  
José Fernando Moreno Gutiérrez  
Luis Fernando Melo Velandia

Núm. 677  
2011

# Borradores de ECONOMÍA



ta - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

# **PRONÓSTICO DE INCUMPLIMIENTOS DE PAGO MEDIANTE MÁQUINAS DE VECTORES DE SOPORTE: UNA APROXIMACIÓN INICIAL A LA GESTIÓN DEL RIESGO DE CRÉDITO\***

**JOSÉ FERNANDO MORENO GUTIÉRREZ**  
(jmorengu@banrep.gov.co)

**LUIS FERNANDO MELO VELANDIA**  
(lmelevel@banrep.gov.co)

BANCO DE LA REPÚBLICA

**RESUMEN.** Este documento describe la metodología desarrollada por Vapnik (1995), denominada máquinas de vectores de soporte (SVM, por sus siglas en inglés) y realiza dos aplicaciones al caso de clasificación de agentes para el otorgamiento de créditos a partir de sus características. El primer caso de estudio clasifica individuos de un banco alemán. En el segundo caso se pronostica el incumplimiento del pago de créditos comerciales otorgados a empresas colombianas utilizando las características iniciales del crédito. SVM se compara con dos metodologías utilizadas en el análisis de este tipo de problemas, regresión logística y análisis lineal discriminante. Los resultados arrojan un mejor desempeño en la predicción por parte de SVM respecto a las otras dos metodologías.

*Palabras claves.* Clasificación, máquinas de aprendizaje, riesgo de crédito, support vector machines.

*Códigos de clasificación JEL.* C44, C50, C60, G32.

---

\* Los resultados y opiniones son responsabilidad exclusiva de los autores y su contenido no compromete al Banco de la República ni a su junta directiva. Los autores agradecen los comentarios y sugerencias de Fernando Arias, Rubén Loaiza y Santiago Téllez. Los autores también agradecen al Departamento de Estabilidad Financiera de la SG-MR del Banco de la República por proporcionar la base de datos del caso colombiano.

## 1. INTRODUCCIÓN

La estimación del riesgo en el que incurren los establecimientos financieros al otorgar un crédito se ha vuelto una labor de suma importancia para su buen funcionamiento. Este riesgo denominado riesgo de crédito, está asociado con la probabilidad de pago de un agente al que se le otorga un crédito. El riesgo de crédito es importante en la medida en que gran parte de la solvencia de las instituciones del sistema financiero depende del pago de los créditos otorgados a los distintos agentes de la economía.

La crisis financiera de 2008 es una muestra de la importancia que tiene la estimación del riesgo de crédito dentro del sistema financiero global. Muchos argumentan que la crisis se presentó debido a la mala administración de este tipo de riesgo puesto que una gran cantidad de préstamos fueron otorgados a personas que no cumplían con los requerimientos mínimos de solvencia para el cumplimiento de sus obligaciones. Por lo tanto, los académicos han venido aplicando y desarrollando nuevas técnicas que permitan determinar el comportamiento de un deudor frente al pago de sus créditos (Thomas, 2009).

A su vez, recientemente la minería de datos y las denominadas máquinas de aprendizaje se han venido desarrollando rápidamente lo que ha permitido su aplicación en varios campos de la ciencia (Xu et al., 2009). Vapnik (1995) presentó un tipo de máquina de aprendizaje, denominada máquina de vectores de soporte (SVM, por sus siglas en inglés), que ha empezado a ser utilizada de forma recurrente en problemas de clasificación.

El problema de estimación del riesgo de crédito puede verse como un problema de clasificación de agentes entre los que pagan y los que no. Por lo tanto, el uso de SVM aplicado al problema de riesgo de crédito ha venido creciendo debido a que esta técnica ha demostrado mayor precisión que las usualmente utilizadas (Yu et al., 2008).

El presente documento describe la metodología de Vapnik (1995), la aplica al problema de clasificación en el riesgo de crédito y compara su desempeño con otras metodologías. Para lo anterior, se realizan dos aplicaciones de la metodología de SVM. En primer lugar, se utiliza una base de datos de préstamos a individuos de un banco alemán y en segundo lugar una base de datos de préstamos comerciales a empresas para el caso colombiano. Luego se compara SVM con las metodologías utilizadas frecuentemente en la solución de estos problemas de clasificación, regresión logística y análisis lineal discriminante (LDA, por sus siglas en inglés).

El documento se divide en seis secciones, siendo esta introducción la primera. La sección dos describe la importancia del riesgo de crédito y presenta una breve comparación entre metodologías aplicadas al riesgo de crédito para la elaboración de *scoring*. En la sección tres se presenta una descripción de una máquina de aprendizaje. Posteriormente, en la sección cuatro se presenta la metodología de Vapnik (1995). En la sección cinco se presentan dos aplicaciones al problema de clasificación en el riesgo de crédito. Finalmente, en la sección seis se realizan algunos comentarios finales.

## 2. RIESGO DE CRÉDITO Y MODELACIÓN

El crédito es un medio de pago de vital importancia para el desarrollo económico, debido a que ha permitido que individuos, empresas y estados logren adquirir recursos para activar el desarrollo de la economía en general. Los individuos lo han utilizado para satisfacer su consumo, mientras que las empresas y los estados para satisfacer necesidades de inversión. El financiamiento, a partir de crédito, ha permitido el desarrollo tecnológico de las empresas y por lo tanto una mejora en las condiciones de vida de los consumidores que, a través de nuevos productos, suplen de mejor manera el conjunto de necesidades.

El uso masivo del crédito como medio de pago al inicio del siglo XX generó implícitamente un riesgo asociado a la posibilidad de incumplimiento en el pago de las obligaciones por parte de los deudores. Este crecimiento acelerado del crédito sumado a la falta de regulación generaron un crecimiento económico basado en crédito con fines especulativos, lo que desencadenó posteriormente la crisis de 1929.

A partir de ese momento, los economistas replantearon el papel que podía jugar la administración y gestión del riesgo crediticio. La modelación del riesgo de crédito adquiere importancia dentro del sistema financiero, por lo que se desarrollan trabajos enfocados a mejorar la administración de este riesgo y por lo tanto el otorgamiento de créditos.

Uno de los primeros trabajos al respecto es desarrollado por Durand (1941) para crédito de consumo, basado en el método de clasificación desarrollado por Fisher (1936). Este fue uno de los primeros trabajos que permitió la cuantificación de la propensión de un individuo a no pagar. Esta valoración del riesgo sustentó, en parte, la introducción de las tarjetas de crédito como medio de pago en 1950. El uso de este novedoso medio de pago desencadenó un *boom* del crédito acompañado de una mayor preocupación de los bancos por no equivocarse en el momento de otorgar créditos (Thomas, 2009).

En la década de los sesentas se calcularon índices del riesgo asociado a la suma de las probabilidades de incumplimiento relacionadas con las características de los individuos. Estos índices, llamados *scores*, fueron la primera herramienta que permitió catalogar a los individuos o empresas de una muestra por medio de un *ranking*. El *score* o puntaje puede definirse como una medida cuantitativa que mide el riesgo basado en las características de los agentes (Gourieroux y Jasiak, 2007) y puede analizarse por medio de metodologías que intentan predecir la posibilidad de que un nuevo deudor no cumpla con la obligación adquirida con sus acreedores.

Por otro lado, con el fuerte crecimiento del mercado financiero a partir de los setentas y las crisis económicas que le siguieron (crisis de deuda en latinoamérica, crisis asiática, crisis *subprime*, entre otras) se estableció un marco regulatorio que llevó a la academia, instituciones financieras y entes reguladores a discutir de manera profunda la medición de los riesgos financieros.<sup>1</sup>

---

<sup>1</sup>Este marco regulatorio se basa actualmente en el acuerdo de Basilea II. Para mayor información sobre este acuerdo y acuerdos anteriores ver Brosnahan y Lee (1989) y Comité de Supervisión Bancaria de Basilea (2004).

Las principales metodologías de medición de riesgo crediticio se basan en métodos estadísticos y matemáticos<sup>2</sup>. Respecto a la construcción de *scores*, los métodos tradicionales que frecuentemente se utilizan son las técnicas de regresión (LM), análisis discriminante (LDA) y árboles de decisión (DT), entre otros. Recientemente, han tomado fuerza nuevas técnicas, dentro de las que se encuentran las de inteligencia artificial para el análisis del riesgo de crédito; entre ellas se tienen las redes neuronales (NN), algoritmos evolutivos (EA), máquinas de vectores de soporte (SVM), entre otras.

Para analizar diferencias entre dichas metodologías, Yu et al. (2008) presentan una revisión extensa de literatura, para el periodo 1970 - 2007, de alrededor de 600 trabajos y de una serie de libros especializados en *scoring* y evaluación de riesgo de crédito. Debido a que cada metodología tiene una estructura diferente, es difícil señalar cuál es la mejor. Teniendo en cuenta esta restricción, estos autores utilizan 32 de estos documentos en los cuales se señalan diferencias explícitas entre los métodos. A partir de esta revisión se establecen cuatro criterios de comparación: exactitud, interpretación, sencillez y flexibilidad. Los resultados obtenidos se encuentran en el Cuadro 1.

CUADRO 1. Comparación de metodologías de riesgo de crédito para la elaboración de *scoring*

Método	Exactitud	Interpretación	Sencillez	Flexibilidad
LDA, LOG y PR	★★	★★★★	★★★★	★
DT	★★	★★★★	★★	★
KNN	★	★★★★	★★★★	★
LP	★	★★★★	★★	★★★★
NN	★★★★	★	★	★
EA	★★	★	★	★
RS	★★	★	★★	★
SVM	★★★★	★★	★	★★★★
Hybrid/ensemble	★★★★	★	★	★★

**Fuente:** Yu et al. (2008). Las siglas de los métodos se encuentran en inglés. LDA: Análisis discriminante lineal. LOG: Regresión logística. PR: Regresión Probit. DT: Árboles de decisión. KNN: *K-nearest neighbors*. LP: Programación Lineal. NN: Redes neuronales. EA: Algoritmos evolutivos. RS: *Rough Sets*. SVM: Máquinas de vectores de soporte. Hybrid/ensemble: Híbridos.

El Cuadro 1 indica que la selección del método más apropiado depende del criterio utilizado. Por ejemplo, si se tiene en cuenta el criterio de flexibilidad, los métodos con mejor desempeño son programación lineal y máquinas de vectores de soporte. Es importante anotar, que el de exactitud se puede considerar como el más importante de estos criterios puesto que la precisión en la predicción del comportamiento de un nuevo cliente es fundamental para una buena gestión de riesgo del crédito.

Bajo este último criterio, dos de los modelos que hacen parte de la familia de máquinas de aprendizaje (SVM y NN), son los que presentan mejor comportamiento. Sin embargo, SVM es mejor

<sup>2</sup>Las distintas metodologías dependen de la forma en que se aborda el problema. En la literatura se encuentra otra serie de modelos que recogen información adicional y que se utilizan para solucionar distintos interrogantes o para complementar los modelos de puntaje. Estos temas pueden ser consultados en Thomas (2009), Bielecki y Rutkowski (2002), Bluhm et al. (2003), Servigny y Renault (2004) McNeil et al. (2005) y Lando (2004), entre otros.

puesto que supera en interpretación y flexibilidad a NN. Además, su uso para gestión del riesgo de crédito a tomado cada vez más relevancia en la literatura especializada en este tema.

**2.1. El problema de clasificación en el riesgo de crédito.** Uno de los principales problemas en la evaluación del riesgo de crédito está relacionado con la capacidad de predicción del incumplimiento de obligaciones financieras de los individuos. Para resolver este problema se debe utilizar toda la información con la que cuentan los prestamistas. Por ejemplo, los bancos cuentan con la información de sus clientes (personas o empresas) o las calificadoras con la información del comportamiento de la economía de los países.

En este documento se tendrá en cuenta el conjunto de variables que afectan el comportamiento crediticio de las personas. Los bancos solicitan a sus clientes información acerca de su edad, nivel de ingresos, estrato socioeconómico, experiencia crediticia, educación, etc; y partir de esta información clasifica a las personas que desean recibir un crédito. Para realizar esta clasificación a partir de la información de cada cliente, el banco debe considerar el comportamiento que ha observado de sus clientes antiguos. Este análisis tiene en cuenta tanto las variables antes mencionadas como el hecho de que cada individuo haya entrado o no en el incumplimiento de sus obligaciones.

Con este conjunto de información, el problema de clasificación de nuevos clientes se puede plantear de la siguiente manera (Yu et al., 2008):

- $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{Mk})^T$  se define como un conjunto de  $M$  variables aleatorias que describen la información del cliente  $k$ , para  $k = 1, 2, \dots, N$
- Para cada individuo  $k$  se tienen la información  $\{\mathbf{x}_k, y_k\}$  con  $k = 1, \dots, N$ , donde  $y_k$  es el resultado observado del pago de la deuda. Por ejemplo,  $y_k = 1$  si cumple con la obligación y  $y_k = -1$  si no la cumple.

Visto de esta forma el problema se resume en realizar una buena clasificación de los clientes; es decir, predecir  $y_k$  a partir de sus características, para un determinado punto del tiempo.

Mediante el modelo básico de búsqueda funcional para máquinas de aprendizaje, el problema se resuelve al encontrar aquella máquina de aprendizaje ( $ML$ ) con la mayor capacidad para imitar o identificar a  $S$ .

### 3. MÁQUINAS DE APRENDIZAJE

En muchas situaciones, los algoritmos computacionales constituyen una forma de solucionar problemas matemáticos complejos. Estos algoritmos son un conjunto de instrucciones, reglas y variables (entradas) bien definidas que, a través de un conjunto de pasos, generan una solución (salida) al problema planteado. Por ejemplo, supongamos que se desea ordenar un conjunto de números en forma ascendente. Para solucionar este problema, se puede plantear un algoritmo donde la entrada son los números y la salida los números ordenados (Alpaydin, 2004).

Es responsabilidad del programador implementar correctamente las especificaciones del problema mediante la construcción de un conjunto de instrucciones que generen una solución eficiente. Sin

embargo, no todos los problemas son iguales por lo que cada uno requiere diferentes nivel de detalle y distintos métodos de solución.

Dado el amplio uso de la informática, estos algoritmos se han convertido en herramientas poderosas para solucionar diferentes tipos de problemas. Por ejemplo, las personas pueden tener acceso a sus cuentas bancarias a través de Internet en cualquier momento; esto es posible por la existencia de un algoritmo que identifica las cuentas bancarias de cada individuo a través de un código único como la cédula o su número de cuenta y una clave de acceso que debe recordar el usuario.

Sin embargo, hay problemas complejos que surgen del mismo crecimiento tecnológico. Por ejemplo, para tener mayor seguridad al acceder a la cuenta, se podría establecer que los clientes deberían utilizar señales de voz introducidas por medio de un micrófono. Este problema se hace bastante complejo de resolver para el programador del algoritmo ya que sería necesario reconocer la voz a partir de los patrones propios de cada cliente.

Este tipo de problemas, en los que no existe un patrón único a seguir, han llevado a replantear un problema filosófico y técnico motivado por la aparición de la robótica en el siglo pasado: brindarles a las máquinas la posibilidad de aprender. Mientras que el problema filosófico es aún motivo de debate, la discusión técnica ha recibido mayor reconocimiento con la llegada de la era electrónica e informática. Se ha demostrado que las máquinas presentan un nivel significativo de habilidad para aprender, a pesar de que la definición de los límites para este proceso están lejos de definirse (Cristianini y Shawe-Taylor, 2000).

Rosenblatt (1958), un psicólogo social de la universidad de Cornell, sugirió una máquina de aprendizaje llamada Perceptron para solucionar un problema simple de la biología: el problema de entender cómo los organismos clasifican y reconocen patrones. Este problema surge al intentar analizar la capacidad que tienen los organismos para generar una percepción sobre otro elemento. Para él, el problema consistía en entender cómo un sistema biológico detecta la información del mundo físico, de qué forma almacena esta información y cómo la usa para reconocer su ambiente y moverse dentro de él. En resumen, entender cómo aprenden los organismos.

Si volvemos al problema de identificación de voz, es posible ver la relación con el Perceptron. Para identificar una voz se requiere buscar una forma mediante la cual una máquina pueda almacenarla, detectarla y reconocerla. Entonces, se debe determinar de qué forma la máquina que recibe las voces, puede aprender de ellas y de esta forma lograr distinguir las.

Otros problemas que pueden abordarse por este tipo de metodologías se han generado con el crecimiento de los sistemas de información, el desarrollo de la medicina, la industria, las telecomunicaciones, el sector financiero, etc. Así, un ejemplo de estos son los establecimientos comerciales y financieros que recogen información sobre sus clientes y el comportamiento que ellos presentan respecto a un producto en particular. Esto ha llevado a que las empresas asocien las características y comportamientos de sus clientes con los productos que ofrecen o compran.

La necesidad de analizar y explotar esta información ha llevado a la utilización de la denominada minería de datos. Se le denomina minería ya que, al igual que en una mina, se extraen elementos

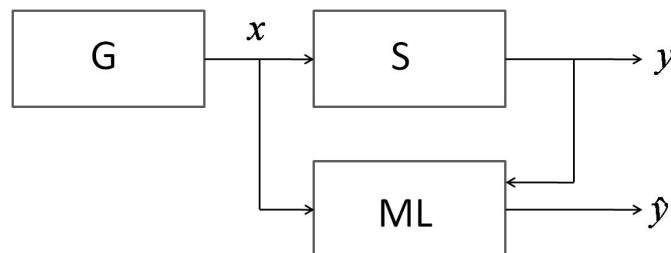
con gran valor que se utilizan para facilitar otros tipos de procesos. Millones de correos electrónicos, compras de productos registrándose segundo a segundo e información sobre las características de los clientes de un banco son ejemplos de estos “minerales”<sup>3</sup>.

Las máquinas de aprendizaje pueden aplicarse sobre estos datos para solucionar diversos tipos de problemas. Para esto, es importante que la máquina sea capaz de aprender de cada evento registrado. Esta capacidad es la misma que tienen los seres humanos; por ejemplo, un niño aprende cuál es el color azul observando una serie de elementos de cualquier color en donde se encuentran una gran cantidad de objetos azules, que son señalados por sus padres. Es necesario, entonces, que se describa el proceso en el que un sistema puede aprender de la experiencia por medio de unas entradas y de una forma funcional que las convierta en salidas. Para poder aprender, las máquinas deben tener información previa que se materializa en ejemplos o muestras de entradas y salidas o eventos ya realizados (Cristianini y Shawe-Taylor, 2000).

Teniendo en cuenta lo anterior, una máquina de aprendizaje se puede definir como aquella máquina que tiene la habilidad para mejorar su propio rendimiento a partir de eventos ya realizados. Para entender esta definición se puede ver el modelo general de aprendizaje que presenta Vapnik (1998)<sup>4</sup>.

En la Figura 1 se considera un modelo de búsqueda funcional donde  $G$  es el generador de datos,  $S$  es el operador del objetivo o supervisor y  $ML$  es la máquina de aprendizaje.

FIGURA 1. Modelo de aprendizaje a partir de ejemplos.



Fuente: Vapnik (1998).

Entonces,  $G$  genera los vectores  $x \in X$  independiente e idénticamente distribuidos (*i.i.d*) de acuerdo a una función de distribución no conocida (pero fija),  $F(X)$ . Estos vectores son las entradas que recoge  $S$ , y posteriormente retorna la salida  $y \in Y$ , para todo vector  $x$ . Este modelo supone que  $S$  existe y no cambia, aunque es desconocido.

Por último,  $ML$  observa  $N$  parejas  $(x_1, y_1), \dots, (x_N, y_N)$  o conjunto de entrenamiento. A partir de este conjunto la máquina construye algún operador el cual es usado para predecir la respuesta del supervisor,  $y_i$  notada como  $\hat{y}_i$ , para un valor de entrada  $x_i$  generado por  $G$ .

<sup>3</sup>Estos datos han permitido estudiar una gran cantidad de problemas en el mercado de valores, en la medición y gestión del riesgo, la detección de fraudes, el control de calidad, el diagnóstico médico y la calidad de las llamadas, etc. Ver Alpaydin (2004) y Hastie et al. (2009).

<sup>4</sup>Esta ilustración y descripción del comportamiento de una máquina de aprendizaje es tomada de Vapnik (1998), páginas 19-21.



Este análisis supone que  $S$  produce la salida  $y$  para el vector  $x$  de acuerdo a la distribución condicional  $F(y|x)$ . Posteriormente,  $ML$  construye una aproximación al operador desconocido de dos formas:

- *Imitar  $S$* : Bajo esta opción se construye un operador que genera la mejor predicción de las salidas de  $S$ .
- *Identificar  $S$* : Se genera un operador que se aproxime a  $S$ .

Por lo tanto, el proceso de aprendizaje selecciona la función más apropiada de un conjunto de funciones dadas. Esto lleva a que una máquina de aprendizaje sea aquella que a partir de la información del pasado (conjunto de entrenamiento) sea capaz de aprender, es decir, encontrar la mejor función.

#### 4. MÁQUINAS DE VECTORES DE SOPORTE<sup>5</sup>

SVM es una máquina de aprendizaje novedosa, desarrollada por Vapnik (1995), que se ha convertido en una metodología importante para solucionar problemas de clasificación. Su buen desempeño ha llevado a su uso en una gran variedad de problemas<sup>6</sup>.

Luego del desarrollo presentado por Vapnik (1998), algunos investigadores como Yu et al. (2010), Bellotti y Crook (2009), Xu et al. (2009), Li et al. (2006), Huang et al. (2004), entre otros, han utilizado SVM para solucionar el problema de evaluación del riesgo de crédito presentado en la sección 2.1 y han mostrado que son mejores en comparación a otras metodologías usuales.

Las principales ventajas de utilizar SVM son<sup>7</sup>:

- Comparada con otras metodologías requiere menos supuestos sobre los datos de entrada. No se asume, por ejemplo, normalidad o continuidad.
- Su desempeño no depende del tamaño o de la dimensión de la muestra. Esto permite que pueda ser utilizado para una cantidad limitada de datos en contraste a otras metodologías que presentan mejor desempeño cuando el tamaño de la muestra es grande.
- Se resuelve mediante programación cuadrática lo que hace que su solución sea única, óptima y global.
- El algoritmo se puede ajustar a problemas no lineales, lo que hace que sea un método generalizable.

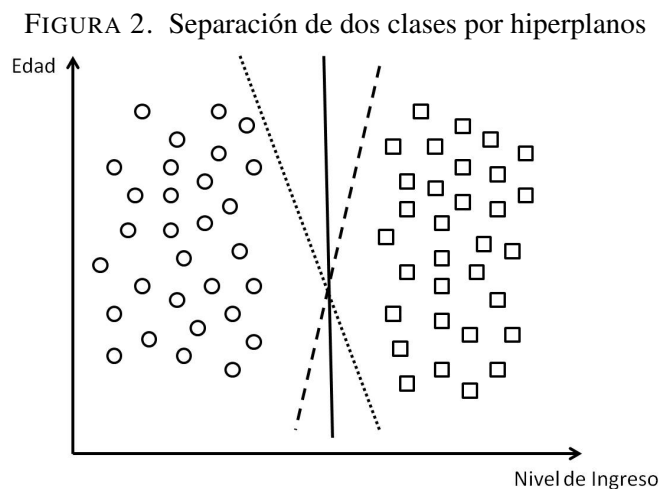
<sup>5</sup>Esta metodología es desarrollada a partir de la teoría del aprendizaje estadístico descrita en Vapnik (1995); ésta se basa en conceptos claves como el principio de minimización del riesgo estructural y la dimensión Vapnik- Chervonenkis, para mayor detalle de estos conceptos ver Cristianini y Shawe-Taylor (2000), Vapnik (1998) y Apéndice A. La descripción de la metodología de SVM que se presenta en este documento se basa principalmente en los trabajos de Vapnik (1998), Burges (1998) y Cristianini y Shawe-Taylor (2000).

<sup>6</sup>Para ejemplos de este tipo de problemas ver Gong y Xu (2007), Preisach et al. (2008), Boursas et al. (2008), Hamel (2009), Liang y Xu (2011), entre otros.

<sup>7</sup>Ver Cristianini y Shawe-Taylor (2000), Li et al. (2006), Kim y Sohn (2010) y Velásquez et al. (2010).

**4.1. Caso linealmente separable.** El caso más sencillo que se puede analizar a partir de SVM es aquel donde el conjunto de entrenamiento es linealmente separable. Este caso es relevante para el análisis de SVM ya que permite entender el conjunto de problemas que se pueden abordar desde esta metodología. Un ejemplo básico de este caso se puede representar a partir de un problema con dos características y una salida.

En el análisis de evaluación de riesgo de crédito las dos características de entrada podrían ser la edad del cliente y su nivel de ingresos y la salida el pago o no pago de sus obligaciones. La Figura 2 muestra la relación hipotética entre estas variables, cada punto es el vector de características de cada cliente en la muestra que además refleja el comportamiento a partir del cumplimiento o incumplimiento de sus obligaciones (cuadrados o círculos, respectivamente).

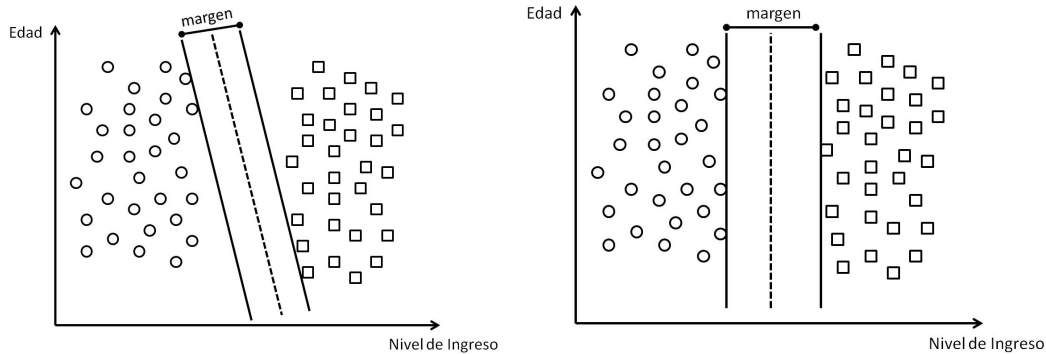


**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. La gráfica muestra la relación entre la edad, el nivel de ingresos y el comportamiento de un conjunto de clientes. Los cuadrados representan a los individuos que han cumplido con sus obligaciones mientras que los círculos el caso contrario. Por medio de líneas se genera una posible clasificación de los puntos según el comportamiento de los clientes.

En la Figura 2 se pretende separar los individuos de acuerdo a su cumplimiento. Se observan múltiples formas en las que se pueden separar las dos clases por medio de un hiperplano, en este caso de una línea. Si se toman dos de los posibles hiperplanos que separan a los clientes tal como se ve en la Figura 3, se puede observar que el margen máximo entre los dos casos está dado por el de la derecha. Sin embargo, no se puede asegurar que aquel margen es el mayor respecto a todos los posibles hiperplanos de separación que pueden encontrarse dentro de este ejemplo. Por lo tanto, el problema consiste en encontrar un hiperplano que maximice el margen o distancia que separa estas clases.

Para formalizar esta idea matemáticamente y encontrar una solución única, se puede recurrir a una formulación que relacione el hiperplano que maximiza el margen de separación con un determinado conjunto de entrenamiento. Por consiguiente, si se cuenta con el conjunto de entrenamiento,  $CE = \{(\mathbf{x}_k, y_k) | \mathbf{x}_k \in \mathbb{R}^N, y_k \in \{-1, 1\}\}_{k=1, \dots, N}$  que contiene las características y el comportamiento de los clientes respecto a sus obligaciones, entonces es posible encontrar un hiperplano que separe a los individuos que han cumplido con sus obligaciones y los que no.

FIGURA 3. Comparación entre dos hiperplanos de separación



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. La gráfica muestra dos posibles hiperplanos de separación para el conjunto de datos. La gráfica de la derecha representa el margen con mayor margen de separación.

Aquellos puntos  $\mathbf{x}$  que se encuentran en el hiperplano de separación satisfacen la siguiente relación:

$$\mathbf{x} \cdot \mathbf{w} + b = 0 \quad (1)$$

Donde  $\mathbf{w}$  es un vector normal y perpendicular al hiperplano y  $\mathbf{x} \cdot \mathbf{w}$  corresponde al producto punto entre los dos vectores.

El margen, por lo tanto, se puede ver como la distancia entre las proyecciones perpendiculares del punto a la izquierda y el de la derecha más cercanos al hiperplano de separación. Para formular lo anterior, es posible suponer que los datos que pertenecen al conjunto de entrenamiento satisfacen las siguientes restricciones<sup>8</sup>:

$$\mathbf{x}_k \cdot \mathbf{w} + b \geq +1 \quad \text{para } y_k = +1 \quad (2)$$

$$\mathbf{x}_k \cdot \mathbf{w} + b \leq -1 \quad \text{para } y_k = -1 \quad (3)$$

Que a su vez se puede expresar como:

$$y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall k \quad (4)$$

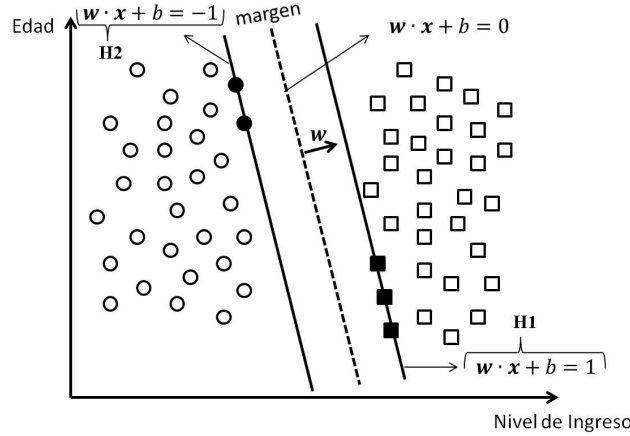
El planteamiento completo del problema considera, adicionalmente, los puntos para los que se cumple la igualdad en (2) y se encuentran en el hiperplano  $H1 : \mathbf{x}_k \cdot \mathbf{w} + b = 1$ . También, se toman los puntos en los que se cumple la igualdad en (3) que se encuentran en el hiperplano  $H2 : \mathbf{x}_k \cdot \mathbf{w} + b = -1$ .

Con base en los conceptos anteriores se puede plantear el problema que encuentra una solución óptima y única para  $\mathbf{w}$  y  $b$ , de tal forma que se maximiza el margen. En el caso bidimensional, la Figura 4 ilustra estos conceptos. Hasta el momento los gráficos realizados involucran hiperplanos

<sup>8</sup>El teorema 4.18 de Cristianini y Shawe-Taylor (2000) involucra el margen geométrico que es el margen de un vector normalizado. De esta forma, es posible optimizar el margen geométrico fijando el margen como 1. Hiperplanos con un margen funcional igual a 1 son conocidos algunas veces como hiperplanos canónicos. El uso de estos hiperplanos canónicos facilita el desarrollo matemático del problema.

en dos dimensiones. En el Apéndice B se encuentran las representaciones gráficas con hiperplanos en tres dimensiones.

FIGURA 4. Problema de clasificación a partir de un hiperplano óptimo



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. La gráfica presenta los conceptos básicos para entender el problema de clasificación a partir de SVM.

Para desarrollar el problema es necesario definir el margen  $m$ . En el Apéndice C se muestra que  $m$  es una función de la norma euclidiana de  $\mathbf{w}$ ,  $\|\mathbf{w}\|$ , tal como sigue:

$$m = \frac{2}{\|\mathbf{w}\|} \tag{5}$$

Entonces, el problema radica en maximizar el margen  $m$  definido en la ecuación (5). Este problema se puede expresar de la siguiente manera:

$$\begin{aligned} &\underset{\mathbf{w}, b}{\text{argmín}} \quad \|\mathbf{w}\|^2 && (6) \\ &\text{s.a.} \quad y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall k \end{aligned}$$

La expresión (6) se puede transformar en su forma dual correspondiente a partir de su lagrangiano asociado:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{k=1}^N \alpha_k [y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1] \tag{7}$$

Donde  $\alpha_k \geq 0$  son los multiplicadores de Lagrange.

Las siguientes condiciones de primer orden asociadas a este problema son:<sup>9</sup>

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = \mathbf{w} - \sum_{k=1}^N y_k \alpha_k \mathbf{x}_k = 0$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = \sum_{k=1}^N y_k \alpha_k = 0$$

Por lo tanto:

$$\mathbf{w} = \sum_k^N y_k \alpha_k \mathbf{x}_k \quad (8)$$

$$0 = \sum_k^N y_k \alpha_k \quad (9)$$

Reemplazando (8) y (9) en (7):

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}) - \sum_{k=1}^N \alpha_k \left[ y_k \left( \mathbf{x}_k \cdot \left( \sum_{l=1}^N y_l \alpha_l \mathbf{x}_l \right) + b \right) - 1 \right] \\ &= \frac{1}{2} \left( \sum_{k=1}^N y_k \alpha_k \mathbf{x}_k \cdot \sum_{l=1}^N y_l \alpha_l \mathbf{x}_l \right) - \sum_{k=1}^N \sum_{l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) - b \sum_k^N y_k \alpha_k + \sum_k^N \alpha_k \\ &= \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) - \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) + \sum_{k=1}^N \alpha_k \\ W(\boldsymbol{\alpha}) &= \sum_k^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) \end{aligned} \quad (10)$$

El cambio de notación de  $L(\mathbf{w}, b, \boldsymbol{\alpha})$  por  $W(\boldsymbol{\alpha})$  refleja la nueva representación del problema. Ahora es necesario encontrar el hiperplano óptimo a partir de los coeficientes  $\alpha_k^*$  que maximizan la ecuación (10). El anterior problema se puede formular como:

$$\begin{aligned} \operatorname{argmáx}_{\alpha_1, \dots, \alpha_N} \quad & \sum_k^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) \\ \text{s.a.} \quad & \sum_k^N y_k \alpha_k = 0, \\ & \alpha_k \geq 0, \quad k = 1, \dots, N. \end{aligned} \quad (11)$$

Si se utilizan los coeficientes  $\alpha_k^*$ ,  $k = 1, \dots, N$ , en la ecuación (8), se tiene que:

$$\mathbf{w}^* = \sum_k^N y_k \alpha_k^* \mathbf{x}_k$$

Es de notar que el valor de  $b$  no está definido en (11), por lo que  $b^*$  debe ser encontrado por medio de (2) y (3) cuando se satisface la igualdad. De esta forma se obtiene:

<sup>9</sup>Nótese que  $\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w}$ .

$$b^* = -\frac{2\mathbf{x}_k \cdot \mathbf{w}^*}{2}$$

Teniendo en cuenta el problema de optimización planteado, la solución óptima para  $\mathbf{w}^*$  y  $b^*$  debe satisfacer las condiciones Karush-Kuhn-Tucker (KKT)<sup>10</sup>:

$$\alpha^*(y_k(\mathbf{x}_k \cdot \mathbf{w}^* + b^*) - 1) = 0, \quad k = 1, \dots, N. \quad (12)$$

Las soluciones no triviales asociadas a la condición (12) se tienen cuando  $\alpha_k^*$  es diferente de cero. Estos valores corresponden a los vectores  $\mathbf{x}_k$ , denominados vectores de soporte, que satisfacen la igualdad:

$$y_k(\mathbf{x}_k \cdot \mathbf{w}^* + b^*) = 1 \quad (13)$$

En la Figura 4 los vectores de soporte son aquellos que forman parte de los hiperplanos  $H1$  y  $H2$  (cuadrados y círculos negros). Estos vectores juegan un rol principal en la construcción del algoritmo en la medida en que el vector  $\mathbf{w}^*$  está definido a partir de ellos:

$$\mathbf{w}^* = \sum_k^N y_k \alpha_k^* \mathbf{x}_k$$

Por lo tanto, el hiperplano óptimo es:

$$f(\mathbf{x}, \boldsymbol{\alpha}^*, b^*) = \sum_{k \in SV}^N y_k \alpha_k^* (\mathbf{x}_k \cdot \mathbf{x}) + b^* \quad (14)$$

Es de notar, que tanto la ecuación (12) como la ecuación (14) no dependen de la dimensión del vector  $\mathbf{x}$  ( $M$ ); éstas dependen del producto punto  $\mathbf{x}_k \cdot \mathbf{x}$ . Como se observará más adelante, este hecho es relevante para la construcción de hiperplanos de separación en espacios de mayor dimensión.

**4.2. Caso linealmente no separable.** En la Figura 5 se observa que la mayoría de clientes que cumplen con sus obligaciones tienen ingresos altos. Sin embargo, existen personas que, aunque presentan un nivel alto de ingresos, incumplen con sus obligaciones. De forma análoga, también hay personas con bajos ingresos que cumplen con estos pagos. Esto complica el problema de separación puesto que las clases no pueden ser separadas por un hiperplano. A diferencia del caso linealmente separable, este tipo de escenarios ocurren con mayor frecuencia en la práctica.

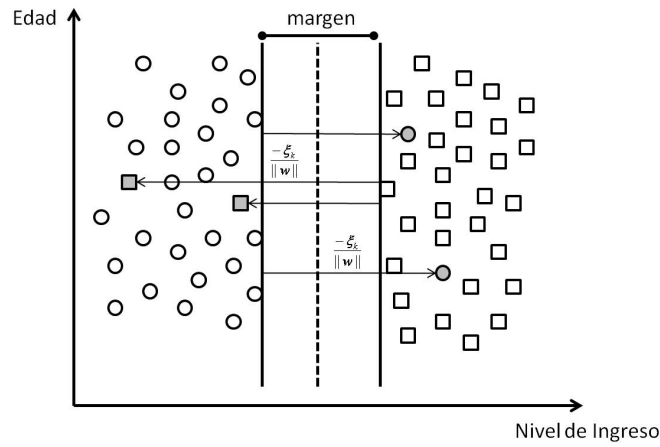
Con el fin de extender la metodología de SVM para casos que no son linealmente separables, se deben relajar las restricciones (2) y (3). Esto se realiza mediante la introducción de una variable positiva  $\xi_k$ ,  $k = 1, \dots, N$ :

$$\mathbf{x}_k \cdot \mathbf{w} + b \geq +1 - \xi_k \quad \text{para } y_k = +1 \quad (15)$$

$$\mathbf{x}_k \cdot \mathbf{w} + b \leq -1 + \xi_k \quad \text{para } y_k = -1 \quad (16)$$

<sup>10</sup>Para mayor detalle véase Vapnik (1998), capítulo 9, sección 9.5.

FIGURA 5. Separación de dos clases por hiperplanos



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. Se presenta el caso en el que los datos no son linealmente separables. Los cuadrados y círculos grises representan la submuestra que impide la separación.

Por lo tanto:

$$y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1 + \xi_k \geq 0, \text{ donde } \xi_k \geq 0 \quad \forall k \quad (17)$$

Este tipo de SVM se denomina máquinas de vectores de soporte con margen suavizado. Este procedimiento penaliza a aquellos puntos que generan la no linealidad; por ejemplo, los cuadrados y círculos grises de la Figura 5.  $\xi_k$  se puede interpretar como una penalización proporcional a la distancia entre estos puntos y los hiperplanos  $H1$  y  $H2$ .

Teniendo en cuenta que el número de vectores mal clasificados debe ser mínimo, el problema descrito en (6) se puede presentar de la siguiente manera:

$$\begin{aligned} \underset{\mathbf{w}, b, \xi}{\text{argmín}} \quad & \|\mathbf{w}\|^2 + C \sum_{k=1}^N \xi_k \\ \text{s.a.} \quad & y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1 + \xi_k \geq 0 \end{aligned} \quad (18)$$

El parámetro  $C$  controla la ponderación entre la penalización de la variable  $\xi_k$  y el tamaño del margen en la función objetivo. A medida que  $C$  aumenta, se penaliza más al error de clasificación.

Este problema se resuelve de la misma manera en que se solucionó el problema (6), por medio del lagrangiano y de su representación dual:

$$L(\mathbf{w}, b, \xi, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{k=1}^N \xi_k - \sum_{k=1}^N \alpha_k [y_k(\mathbf{x}_k \cdot \mathbf{w} + b) - 1 + \xi_k] - \sum_{k=1}^N \mu_k \xi_k \quad (19)$$

Las condiciones de primer orden de (19) son las siguientes:

$$\begin{aligned}\frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha})}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{k=1}^N y_k \alpha_k \mathbf{x}_k = 0 \\ \frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha})}{\partial b} &= \sum_{k=1}^N y_k \alpha_k = 0 \\ \frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha})}{\partial \xi_k} &= C - \alpha_k - \mu_k = 0\end{aligned}$$

Por lo tanto:

$$\mathbf{w} = \sum_{k=1}^N y_k \alpha_k \mathbf{x}_k \quad (20)$$

$$0 = \sum_{k=1}^N y_k \alpha_k \quad (21)$$

$$C = \alpha_k + \mu_k \quad (22)$$

Reemplazando (20), (21) y (22) en (19):

$$\begin{aligned}L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}) &= \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) + C \sum_{k=1}^N \xi_k - \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) - b \sum_{k=1}^N \alpha_k y_k \\ &\quad + \sum_{k=1}^N \alpha_k - \sum_{k=1}^N \alpha_k \xi_k - \sum_{k=1}^N \mu_k \xi_k \\ &= \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) + (C - \mu_k - \alpha_k) \sum_{k=1}^N \xi_k \\ W(\boldsymbol{\alpha}) &= \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l)\end{aligned} \quad (23)$$

Ahora el problema es similar a (11) bajo otras restricciones:

$$\begin{aligned}\operatorname{argm\acute{a}x}_{\alpha_1, \dots, \alpha_N} \quad & \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N y_k y_l \alpha_k \alpha_l (\mathbf{x}_k \cdot \mathbf{x}_l) \\ \text{s.a.} \quad & \sum_{k=1}^N y_k \alpha_k = 0, \\ & 0 \leq \alpha_k \leq C, \quad k = 1, \dots, N.\end{aligned} \quad (24)$$



La solución es de nuevo dada por:

$$\mathbf{w}^* = \sum_{k \in SV}^N y_k \alpha_k^* \mathbf{x}_k$$

Por lo tanto, la diferencia entre el caso linealmente separable y el no linealmente separable, es que ahora  $\alpha_k$  tiene un límite superior  $C$ . A partir de las condiciones KKT<sup>11</sup> es posible encontrar  $b^*$ , además, deducir que  $\xi_k = 0$  para todo  $\alpha_k < C$ .

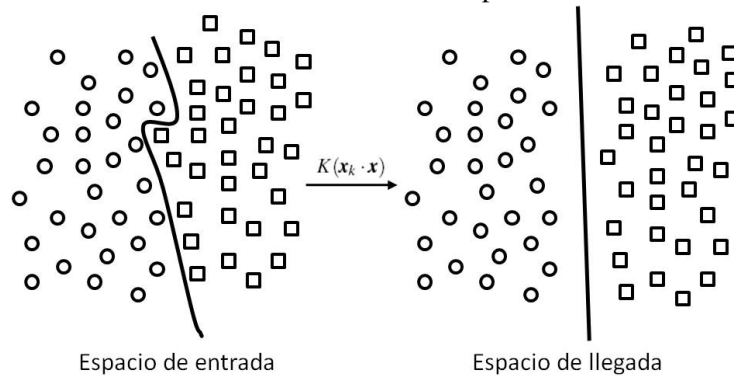
**4.3. SVM no lineales.** Cuando (4) y (17) no son funciones lineales, es decir, es imposible separar linealmente los datos en su espacio de entrada, es necesario emplear un método que permita generalizar el uso de SVM. Antes de describir el procedimiento, vale recordar que los problemas de optimización anteriormente discutidos, dependen únicamente de los datos a través del producto punto  $\mathbf{x}_k \cdot \mathbf{x}$ .

La metodología para generalizar SVM se basa en un procedimiento denominado *kernel trick* (Aizerman et al., 1964), que consiste en asignar a los datos otro espacio euclidiano  $\mathcal{H}$  en el que aproximadamente se tiene una estructura lineal<sup>12</sup>. Sea  $\Phi$  una función tal que:

$$\Phi : \mathcal{X}^n \mapsto \mathcal{H}$$

Ahora el algoritmo depende de los datos por medio del producto punto en el espacio  $\mathcal{H}$ ; es decir, de las funciones de la forma  $K(\mathbf{x}_k \cdot \mathbf{x}) = \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x})$ . Si se utiliza esta función  $K$ , denominada función *kernel*, no será necesario especificar a  $\Phi$ . Por lo tanto, el caso no lineal es equivalente a aplicar una función a los datos en el espacio de entrada. Posteriormente, el algoritmo de aprendizaje es utilizado en el espacio de llegada de la función (Figura 6).

FIGURA 6. Transformación de los datos por medio de un *kernel*



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. La gráfica ilustra la idea que hay detrás de la utilización de un *kernel* para volver linealmente separables los datos en un espacio de llegada generado por la función.

<sup>11</sup>Ver Burges (1998).

<sup>12</sup>Para profundizar en la utilización de *kernels* en SVM ver Schölkopf y Smola (2001) y Cristianini y Shawe-Taylor (2004).

Esta metodología permite encontrar vectores de soporte en el espacio de llegada  $\mathcal{H}$  sin importar el algoritmo utilizado. Por consiguiente, todas las consideraciones hechas en las secciones 4.1 y 4.2 aplican en la medida en que se realiza una separación lineal, pero en un espacio diferente. Por ejemplo, si se tiene en cuenta el caso linealmente separable presentado en la sección 4.1, el hiperplano óptimo para SVM no lineales está dado por:

$$f(\mathbf{x}, \boldsymbol{\alpha}^*, b^*) = \sum_{k \in SV}^N y_k \alpha_k^* K(\mathbf{x}_k, \mathbf{x}) + b^*$$

Sin embargo, para que este procedimiento pueda llevarse a cabo es necesario que la función *kernel* utilizada satisfaga las condiciones de Mercer (Vapnik, 1995):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_k^{\infty} a_k \Phi_k(\mathbf{x}_i) \Phi_k(\mathbf{x}_j), \quad a_k \geq 0$$

y

$$\int \int K(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_i) g(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j > 0$$

Las funciones *kernel* utilizadas frecuentemente en el reconocimiento de patrones por medio de SVM son<sup>13</sup>:

- **Kernel lineal:** Este es un clasificador lineal. Generalmente es utilizado para probar la existencia de no linealidades en el conjunto de entrenamiento. Además, el uso de este *kernel* se recomienda cuando hay vectores de datos dispersos:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \tag{25}$$

- **Kernel polinomial:** Este es un método eficiente y simple para modelar relaciones no lineales. Tiene como desventaja que conforme aumenta el grado del polinomio ( $d$ ) la superficie de clasificación se hace más compleja:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i \cdot \mathbf{x}_j)^d \tag{26}$$

- **Kernel RBF:** *Radial Basis Function* (RBF) es uno de los *kernels* más utilizados en la literatura. Se recomienda utilizarlo cuando no se cuenta con información a priori de los datos. El parámetro  $\sigma$  controla la forma del hiperplano de separación y éste puede optimizarse a partir de métodos de validación cruzada:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{27}$$

<sup>13</sup>Sin embargo, existen multiples especificaciones, algunas de éstas se encuentran en el Apéndice D.

## 5. APLICACIÓN AL PROBLEMA DE RIESGO DE CRÉDITO

En esta sección se muestran dos aplicaciones de la metodología SVM a la clasificación de nuevos clientes financieros dentro del contexto de riesgo de crédito de préstamos bancarios.

**5.1. Datos.** En primer lugar, se utiliza la base de datos de Frank y Asuncion (2010) tomada del repositorio UCI para máquinas de aprendizaje<sup>14</sup>, que contiene información financiera y algunas características de los clientes de un banco alemán.

En segundo lugar, se toma una muestra de datos de la información relacionada con el Formato 341 de la Superintendencia Financiera de Colombia, que contiene información relevante para el otorgamiento de créditos comerciales de empresas colombianas para el tercer trimestre de 2007. Estos datos fueron proporcionados por el Departamento de Estabilidad Financiera y el Departamento de Información Económica del Banco de la República.

Las variables utilizadas en la base de datos de Frank y Asuncion (2010) reflejan características de los clientes referentes a su vida crediticia y otras variables pertinentes para el análisis, como la edad, el estado civil y el género. Por otra parte, las variables utilizadas para el caso colombiano son características propias del crédito como la tasa de interés del préstamo y el monto de la deuda que se adquiere.

La variable que se utiliza para determinar el cumplimiento o no de la deuda, en las dos bases, es el estado de la cuenta que se toma como una variable dicotómica que toma el valor de cumplimiento o incumplimiento. La base de datos alemana tiene 1000 individuos cuya proporción de clientes que no cumplen el pago de la deuda es del 30%. Para el caso colombiano, la muestra cuenta con 3898 empresas, de las cuales tan sólo el 5.79% incumple. El Apéndice E presenta la descripción de las series que conforman estas dos bases de datos.

**5.2. Metodología.** Con el fin de evaluar el desempeño de SVM, se toman dos metodologías utilizadas dentro del contexto de medición del riesgo de crédito a partir de la clasificación de agentes: regresión logística y análisis lineal discriminante (LDA).

La metodología utilizada para la estimación mediante SVM se lleva a cabo en los siguientes pasos:

- i. Cada una de las bases de datos consideradas se divide en muestra de entrenamiento y muestra de pronóstico. Para esta división se toma aleatoriamente el 80% de los datos como muestra de entrenamiento y el restante 20% como muestra de pronóstico.
- ii. Debido a que los datos no son linealmente separables en su espacio de entrada, para la estimación de SVM se utiliza el algoritmo presentado en la sección 4.3<sup>15</sup> y se transforman los datos mediante los *kernels* RBF, polinomial y lineal.

---

<sup>14</sup>Para mayor información sobre el repositorio UC Irvine para máquinas de aprendizaje ver <http://archive.ics.uci.edu/ml/datasets.html>

<sup>15</sup>Posterior al trabajo de Vapnik (1998) se han desarrollado una serie de algoritmos que mejoran el rendimiento de SVM. Sin embargo, los algoritmos que se utilizan en este documento son de uso estándar en la literatura de riesgo de crédito. Otros algoritmos se pueden encontrar en Schölkopf et al. (2001), Schölkopf et al. (2000), Crammer y Singer (2002) y Hsu y Lin (2002), entre otros

- iii. La calibración de los parámetros de los *kernels*<sup>16</sup> se lleva a cabo mediante búsqueda directa. Se selecciona aquella combinación de parámetros que genera el menor error *cross-validation*. La estimación de este error es a través de *K-fold cross-validation*<sup>17</sup>.
- iv. Adicionalmente, utilizando como insumo los parámetros obtenidos en el paso anterior, se utiliza el método numérico de minimización de funciones de Nelder y Mead (1965) para calibrar los parámetros finalmente utilizados<sup>18</sup>.
- v. Luego de la selección de parámetros, se establece el umbral de clasificación según la probabilidad que asigna SVM de estar en una u otra clase. Este umbral se establece sobre la muestra de entrenamiento y luego se utiliza en la clasificación con la muestra de pronóstico. Cabe resaltar que, debido a que la selección del umbral se hace para mejorar el resultado en la muestra de entrenamiento, ésta no siempre genera un mejor resultado en la muestra de pronóstico.
- vi. Por último, se estima SVM con los parámetros seleccionados, se clasifica la muestra de pronóstico y se compara con los valores observados. A partir de esta comparación se realiza la validación del modelo que establece el poder de pronóstico de SVM para nuevos clientes.
- vii. Para validar el modelo se utilizan medidas que cuantifican el desempeño del pronóstico en las dos muestras, tales como:

$$\text{Precisión en la clasificación de buenos clientes (PBC)} = \frac{BB}{BB + BM} \times 100\%$$

$$\text{Precisión en la clasificación de malos clientes (PMC)} = \frac{MM}{MM + MB} \times 100\%$$

$$\text{Precisión en la clasificación total (PTC)} = \frac{BB + MM}{BB + BM + MB + MM} \times 100\%$$

Donde *BB* corresponde a la cantidad de buenos clientes clasificados como buenos, *BM* a la cantidad de buenos clientes clasificados como malos, *MM* a la cantidad de malos clientes clasificados como malos y *MB* a la cantidad de malos clientes clasificados como buenos.

La Figura 7 resume la metodología que se utilizó para la estimación de SVM<sup>19</sup>.

Para el modelo de regresión logística se seleccionan las variables mediante la metodología de regresión *stepwise* (Efroymson, 1960) bajo el criterio de información de Akaike. Para mejorar la predicción se utiliza un umbral al igual que en SVM. También se selecciona otro modelo que cumpla los supuestos y presente un buen pronóstico. Para el caso de LDA se utiliza una prior con la proporción de la muestra utilizada y también se genera un umbral.

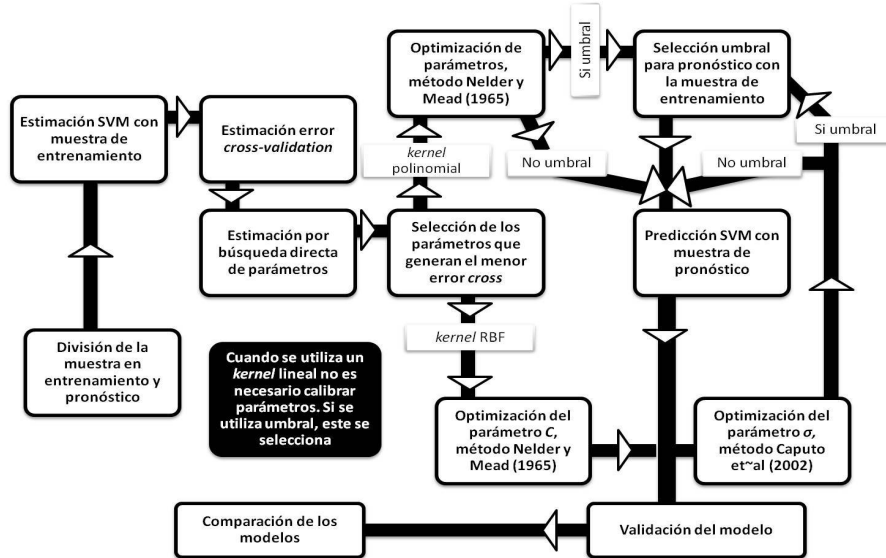
<sup>16</sup>Para el *kernel* lineal no es necesario calibrar ningún parámetro.

<sup>17</sup>La metodología *K-fold cross-validation* consiste en dividir la muestra en *K* partes aleatoriamente; luego se toma cada una de las divisiones como muestras de pronóstico y con las restantes *K* - 1 se estima SVM. Finalmente, el error de *cross-validation* se calcula como el promedio de los *K* errores de predicción de esta muestra. Esta metodología se emplea debido a los posibles problemas de sobreparametrización que se presentan cuando se tiene en cuenta el error sobre la muestra de entrenamiento.

<sup>18</sup>Para el caso de RBF únicamente se calibró el parámetro *C* del problema (18) mediante Nelder y Mead (1965). El parámetro  $\sigma$  de (27) se calibró teniendo en cuenta el método de Caputo et al. (2002) en donde el valor óptimo del parámetro se encuentra entre el cuantil 0,1 y 0,9 de  $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ . Este procedimiento utiliza una muestra del conjunto de entrenamiento y devuelve un vector que contiene los valores de los cuantiles; para este caso se seleccionó el cuantil 0,5

<sup>19</sup>Para ver otras metodologías de estimación de SVM ver Yu et al. (2008) y Huang et al. (2007).

FIGURA 7. Metodología de estimación SVM



Fuente: Gráfica ilustrativa realizada por los autores.

Finalmente, se comparan las metodologías utilizadas dependiendo de *PBC*, *PMC* y *PTC*. Adicionalmente, se tiene en cuenta la curva ROC y su área bajo la curva (AUC, por sus siglas en inglés).

**5.3. Resultados.** Los resultados obtenidos varían dependiendo de cada una de las dos bases de datos anteriormente presentadas. Sin embargo, cabe destacar que para estas bases de datos SVM tiene una buena capacidad de pronóstico respecto a la regresión logística y LDA. Otro resultado importante, es que aquellos modelos que presentan un mejor rendimiento dentro de la muestra de entrenamiento, son los de mayor precisión dentro de la muestra de pronóstico.

Como criterio de comparación, es importante tener en cuenta que clasificar mal a los buenos clientes no es equivalente a clasificar mal a los clientes que cometen default. Generalmente, cometer el segundo error se considera más grave en el sentido en que los clientes con mal comportamiento son aquellos que generan pérdidas reales a los bancos. Por lo tanto, un criterio que debe ser importante a la hora de seleccionar el mejor modelo debe ser la *PMC*, que nos da una idea de cuantos agentes con mal comportamiento crediticio son bien clasificados.

Para el caso de los datos alemanes, como se observa en el Cuadro 2, todos los modelos SVM de la muestra de entrenamiento muestran un mejor desempeño según los criterios *PTC* y la *PBC*. Sin embargo, bajo el criterio *PMC* el modelo LDA presenta mejor desempeño que algunos modelos SVM. Por otro lado, el mejor modelo para la muestra de pronóstico es el SVM que utiliza el *kernel* RBF con umbral. Este modelo presenta los mayores criterios *PTC*, *PMC* y *AUC*.

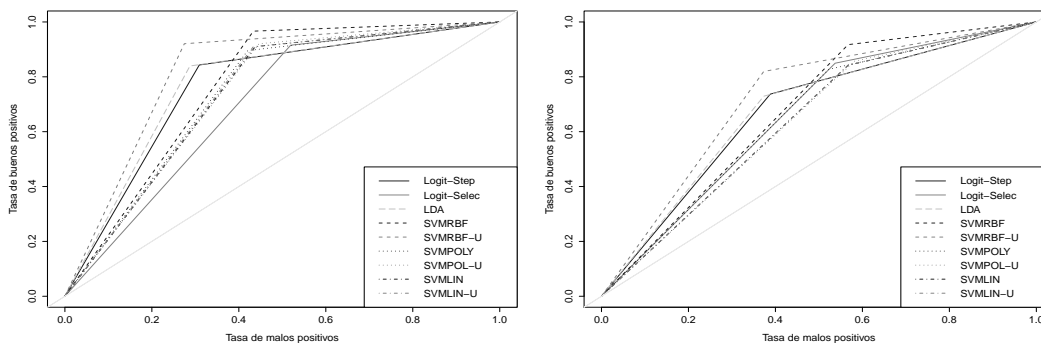
CUADRO 2. Resultados para las muestras de entrenamiento y pronóstico de la base de datos de Frank y Asuncion (2010)

	Muestra de Entrenamiento					Muestra de Pronóstico			
	PTC	PBC	PMC	AUC		PTC	PBC	PMC	AUC
Logit-Step	79.88	84.30	69.10	76.70		69.50	73.68	61.19	67.44
Logit-Selecc	78.80	91.53	48.07	69.80		72.00	84.96	46.27	65.62
LDA	80.25	83.95	71.24	77.59		69.50	72.93	62.68	67.80
SVMRBF	85.00	96.65	56.65	76.65		75.50	91.73	43.28	67.51
SVMRBF-U	86.38	92.06	72.53	82.30		75.50	81.95	62.69	72.32
SVMPOL	80.38	89.59	57.94	73.77		71.00	82.71	47.76	65.23
SVMPOL-U	81.25	92.06	54.94	73.50		71.00	86.47	40.30	63.38
SVMLIN	80.88	91.01	56.22	73.61		70.50	100.00	0.00	63.75
SVMLIN-U	81.00	91.01	56.65	73.83		71.00	84.96	43.28	64.12

Logit-Step corresponde al modelo de regresión logística seleccionado mediante el método *stepwise* de Efron y Tibshirani (1993). Logit-Selecc corresponde al modelo de regresión logística seleccionado con *stepwise* utilizando errores robustos. LDA corresponde al modelo de análisis lineal discriminante. Finalmente, SVMRBF, SVMRBF-U, SVMPOL, SVMPOL-U, SVMLIN y SVMLIN-U corresponden a las máquinas de vectores de soporte con *kernel* RBF sin umbral y con umbral, con *kernel* polinomial sin umbral y con umbral y con *kernel* lineal sin umbral y con umbral, respectivamente.

La curva ROC, Figura 8, confirma los anteriores resultados. Además, cabe resaltar que para este conjunto de datos SVM con *kernel* lineal no presenta un buen comportamiento, lo que muestra evidencia de que los datos presentan no linealidades.

FIGURA 8. Curva ROC modelos estimados para muestra de entrenamiento y pronóstico de la base de datos de Frank y Asuncion (2010)



Las convenciones de esta figura corresponden a las utilizadas en el Cuadro 2.

Los resultados del caso colombiano se presentan en el Cuadro 3 y la Figura 9. Aunque los modelos con mejor desempeño corresponden a SVM, el mejor modelo depende del criterio que se utilice.

En términos generales, la diferencia entre el desempeño de los modelos de regresión logística y el modelo LDA respecto a los SVM no es tan grande como en el caso de los datos alemanes.

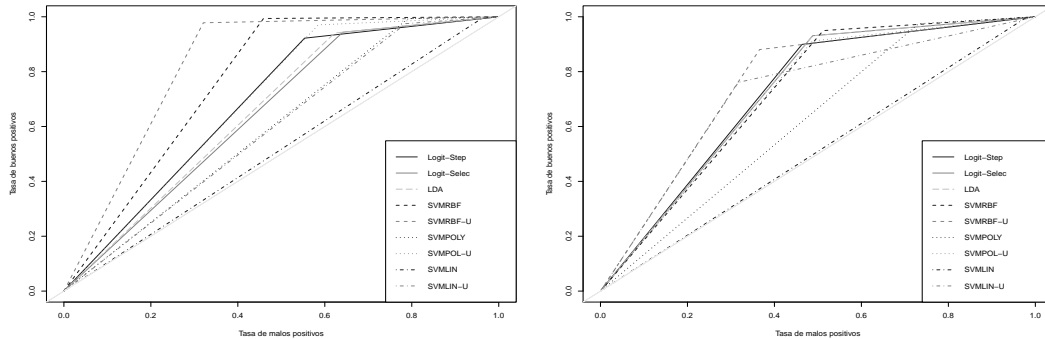
Cabe resaltar que el modelo SVM con *kernel* lineal utilizando umbral tiene el mejor desempeño bajo el criterio PMC. Por otro lado, el modelo SVM *kernel* RBF con umbral presenta el mejor resultado en cuanto al criterio AUC y el segundo mejor respecto a PMC. Estos dos criterios son muy relevantes en la medida en que representan una menor pérdida tangible para la contraparte del crédito.

CUADRO 3. Resultados muestra de entrenamiento y pronóstico de la base de datos Formato 341 Superintendencia Financiera de Colombia

	Muestra de Entrenamiento					Muestra de Pronóstico			
	PTC	PBC	PMC	AUC		PTC	PBC	PMC	AUC
Logit-Step	82.75	92.19	44.65	68.42		82.50	89.93	53.65	71.79
Logit-Selec	82.75	92.19	35.84	65.11		84.50	93.08	51.21	72.15
LDA	82.87	94.07	37.73	65.90		84.50	93.08	51.21	72.15
SVMRBF	90.37	99.37	54.08	76.73		85.50	94.96	48.78	71.87
SVMRBF-U	91.87	97.81	67.92	82.87		83.00	88.05	63.41	75.73
SVMPOL	83.75	99.37	20.75	60.06		83.00	97.48	26.82	62.15
SVMPOL-U	86.00	97.03	41.50	69.27		83.00	91.19	51.21	71.20
SVMLIN	80.50	99.53	3.77	51.65		79.50	100.000	0.00	50.90
SVMLIN-U	82.37	97.50	21.38	59.44		74.50	76.10	68.29	72.19

Las convenciones de este cuadro corresponden a las utilizadas en el Cuadro 2.

FIGURA 9. Curva ROC modelos estimados para muestra de entrenamiento y pronóstico de la base de datos Formato 341 Superintendencia Financiera de Colombia



Las convenciones de esta figura corresponden a las utilizadas en el Cuadro 2.

## 6. COMENTARIOS FINALES

En este documento se presentó la metodología desarrollada por Vapnik (1995) que constituye una alternativa importante y recientemente explorada en la solución de problemas en áreas relacionadas con economía y finanzas. Explícitamente, se utilizó SVM para el problema de clasificación de nuevos clientes bancarios al momento de solicitar un crédito.

La aplicación se realizó para dos tipos de clientes: personas y empresas. Para los primeros se utilizó una base de datos alemana tomada de Frank y Asuncion (2010) y para los segundos se tomó información correspondiente al Formato 341 de la Superintendencia Financiera de Colombia, proporcionada por el Departamento de Estabilidad Financiera y el Departamento de Información Económica del Banco de la República.

La metodología SVM fue comparada con algunos modelos generalmente utilizados en este tipo de análisis: regresión logística y análisis discriminante lineal. Los resultados obtenidos muestran una mayor capacidad de pronóstico de los modelos SVM respecto a los modelos tradicionales. En el caso alemán, el modelo con mejor desempeño fue el SVM con *kernel* RBF utilizando umbral. Para el caso colombiano los mejores modelos fueron SVM con *kernel* lineal y *kernel* RBF, ambos con umbral.

Como extensiones de este trabajo, se podrían utilizar algoritmos híbridos que permitan mejorar la estimación de los parámetros de estos modelos. Por otro lado, los modelos SVM también podrían ser utilizados para obtener pronósticos de series económicas de interés.

## REFERENCIAS

- AIZERMAN, A., E. M. BRAVERMAN, Y L. I. ROZONER (1964): "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, 25, 821–837.
- ALPAYDIN, E. (2004): *Introduction to machine learning*, Adaptive computation and machine learning. MIT Press.
- BELLOTTI, T., Y J. CROOK (2009): "Support vector machines for credit scoring and discovery of significant features.," *Expert Systems with Applications: An international Journal*, 36(2), 3302–3308.
- BIELECKI, T., Y M. RUTKOWSKI (2002): *Credit risk: modeling, valuation and hedging*, Springer finance. Springer.
- BLUHM, C., L. OVERBECK, Y C. WAGNER (2003): *An introduction to credit risk modeling*, Chapman - Hall/CRC financial mathematics series. Chapman - Hall/CRC.
- BOURSAS, L., M. CARLSON, Y W. HOMMEL (2008): *Systems and virtualization management: standards and new technologies : second international workshop, SVM 2008, Munich, Germany, October 21-22, 2008, proceedings*, Communications in computer and information science. Springer.
- BROSNAHAN, M., Y T. C. LEE (1989): "International convergence of capital measurement and capital standards for banks," *Reserve Bank of New Zealand Bulletin*, 52, 1.
- BURGES, C. J. (1998): "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, 2, 121–167.



- CAPUTO, R., K. SIM, F. FURESJO, y A. SMOLA (2002): "Appearance-based Object Recognition using SVMs: Which Kernel Should I Use?," en *Proceedings of NIPS Workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision*. Whistle.
- COMITÉ DE SUPERVISIÓN BANCARIA DE BASILEA (2004): *Convergencia internacional de medidas y normas de capital: Marco revisado* Banco Internacional de Pagos.
- CRAMMER, K., y Y. SINGER (2002): "On the algorithmic implementation of multiclass kernel-based vector machines," *The Journal of Machine Learning Research*, 2, 265–292.
- CRISTIANINI, N., y J. SHAWE-TAYLOR (2000): *An introduction to support vector machines: and other kernel-based learning methods*. Cambridge University Press.
- (2004): *Kernel Methods for Pattern Analysis*. Cambridge University Press.
- DURAND, D. (1941): *Risk Elements in Consumer Installment Financing*. National Bureau of Economic Research.
- EFRONYMSON (1960): "Multiple regression analysis," en *Mathematical Methods for Digital Computers*, ed. por A. Ralston, y H. Wilf. Wiley.
- FISHER, R. (1936): "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7, 179–188.
- FRANK, A., y A. ASUNCION (2010): "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences.
- GONG, Y., y W. XU (2007): *Machine learning for multimedia content analysis*, Multimedia systems and applications series. Springer.
- GOURIEROUX, C., y J. JASIAK (2007): *The Econometric of Individual Risk*. Princeton university press.
- HAMEL, L. (2009): *Knowledge discovery with support vector machines*, Wiley series on methods and applications in data mining. John Wiley & Sons.
- HASTIE, T., R. TIBSHIRANI, y J. FRIEDMAN (2009): *The elements of statistical learning: data mining, inference, and prediction*, Springer series in statistics. Springer.
- HSU, C.-W., y C.-J. LIN (2002): "A Simple Decomposition Method for Support Vector Machines," *Machine Learning*, 46, 291–314.
- HUANG, C.-L., M.-C. CHEN, y C.-J. WANG (2007): "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications: An international Journal*, 33, 847–856.
- HUANG, Z., H. CHEN, C.-J. HSU, W.-H. CHEN, y S. WU (2004): "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, 37, 543–558.
- IVANCIUC, O. (2007): "Applications of Support Vector Machines in Chemistry," en *Reviews in Computational Chemistry*, ed. por K. Lipkowitz, y T. Cundari, vol. 23, pp. 291–400. Wiley.
- KIM, H. S., y S. Y. SOHN (2010): "Support vector machines for default prediction of SMEs based on technology credit," *European Journal of Operational Research*, 201(3), 838–846.
- LANDO, D. (2004): *Credit risk modeling: theory and applications*, Princeton series in finance. Princeton University Press.
- LI, S.-T., W. SHIUE, y M.-H. HUANG (2006): "The evaluation of consumer loans using support vector machines," *Expert Systems with Applications: An international Journal*, 30, 772–782.
- LIANG, Y., y Q. XU (2011): *Support Vector Machines and Their Application in Chemistry and Biotechnology*. Taylor & Francis.
- MCNEIL, A. J., R. FREY, y P. EMBRECHTS (2005): *Quantitative risk management: concepts, techniques and tools*. Princeton Series in Finance.
- NELDER, J. A., y R. MEAD (1965): "A Simplex Method for Function Minimization," *The Computer Journal*, 7(4), 308–313.

- PREISACH, C., H. BURKHARDT, L. SCHMIDT-THIEME, Y R. DECKER (2008): *Data analysis, machine learning and applications: proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7-9, 2007*, Series: Studies in Classification, Data Analysis, and Knowledge Organization. Springer.
- ROSENBLATT, F. (1958): "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, 65, 386–408.
- SCHÖLKOPF, B., J. C. PLATT, J. SHAWE-TAYLOR, A. J. SMOLA, Y R. C. WILLIAMSON (2001): "Estimating the Support of a High-Dimensional Distribution," *Neural Computation*, 13(7), 1443–1471.
- SCHÖLKOPF, B., Y A. J. SMOLA (2001): *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. The MIT Press, 1st edn.
- SCHÖLKOPF, B., A. J. SMOLA, R. C. WILLIAMSON, Y P. L. BARTLETT (2000): "New Support Vector Algorithms," *Neural Computation*, 12, 1207–1245.
- SERVIGNY, A., Y O. RENAULT (2004): *Measuring and managing credit risk*, Standard and Poor's Press. McGraw-Hill.
- THOMAS, L. (2009): *Consumer credit models: pricing, profit, and portfolios*. Oxford University Press.
- VAPNIK, V. (1995): *The Nature of Statistical Learning Theory*. Springer, New York.
- VAPNIK, V. (1998): *Statistical learning theory*, Adaptive and learning systems for signal processing, communications, and control. Wiley.
- VAPNIK, V., Y A. CHERVONENKIS (1971): "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," *Theory of Probability and its Applications*, 16(2), 264–280.
- VELÁSQUEZ, J. D., Y. OLAYA, Y C. J. FRANCO (2010): "Predicción de series temporales usando máquinas de vectores de soporte," *Ingeniare. Revista chilena de ingeniería*, 18, 64 – 75.
- XU, X., C. ZHOU, Y Z. WANG (2009): "Credit scoring algorithm based on link analysis ranking with support vector machine," *Expert Systems with Applications: An international Journal*, 36(2), 2625–2632.
- YU, L., S. WANG, K. LAI, Y L. ZHOU (2008): *Bio-inspired credit risk analysis: computational intelligence with support vector machines*. Springer.
- YU, L., W. YUE, S. WANG, Y K. K. LAI (2010): "Support vector machine based multiagent ensemble learning for credit risk evaluation," *Expert Systems with Applications: An international Journal*, 37(2), 1351–1360.

#### APÉNDICE A. DIMENSIÓN VAPNIK-CHERVONENKIS (VAPNIK Y CHERVONENKIS, 1971)

Este es un concepto clave para el desarrollo de la teoría del aprendizaje estadístico. Es importante en la medida en que su definición permite generalizar el proceso de aprendizaje y solucionar el problema de *overfitting* cuando se utiliza un conjunto finito de datos de entrenamiento y se aplica el principio de minimización del riesgo empírico.

La dimensión VC de un conjunto de funciones indicadoras  $Q(x, \alpha)$ ,  $\alpha \in \Lambda$ , se define como el máximo número  $h$  de vectores  $x_1, \dots, x_h$  que pueden ser separados en dos clases para todas las  $2^h$  posibles formas de usar las funciones del conjunto. Por lo tanto, esta dimensión corresponde al máximo número de vectores que pueden ser separados por estas funciones (Vapnik y Chervonenkis, 1971).

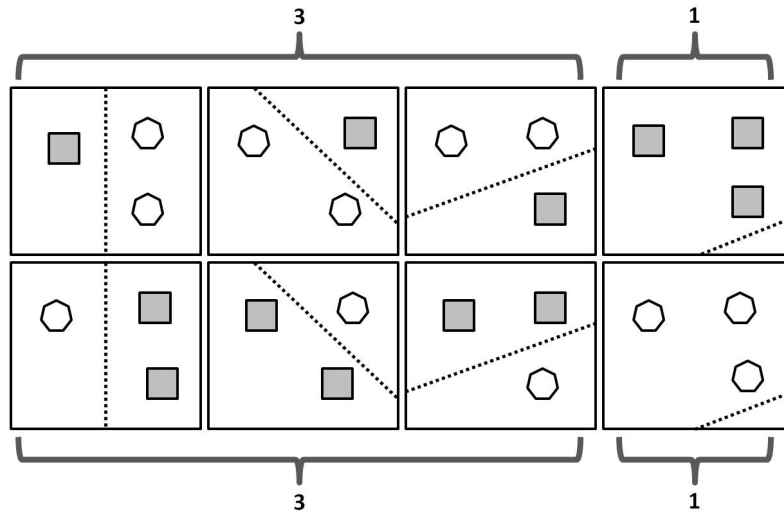
Por lo tanto, para estimar la dimensión VC de un conjunto de funciones  $Q(x, \alpha)$  es suficiente señalar el máximo número de  $h$  vectores  $x_1^*, \dots, x_\ell^*$  que puede ser separado por el conjunto de funciones. Por ejemplo, si se quiere clasificar 3 vectores en dos clases, el número de posibles combinaciones está dado por:

$$2^3 = \sum_{n=0}^3 \binom{3}{n} = 8$$

Si el conjunto de vectores puede ser separado de tal forma por el conjunto de funciones, entonces la dimensión VC es igual al número de vectores separados. Las posibles combinaciones para dos clases se pueden generalizar a partir del triángulo de Pascal, dependiendo del número de elementos del conjunto de vectores.

Para hallar la dimensión VC se debe comprobar que se cumpla con todas las posibles combinaciones para determinado número de elementos dentro del conjunto de vectores. Para el caso de funciones lineales, por ejemplo, si se toman tres vectores estos pueden ser separados bajo todas las posibles combinaciones como se ve en la Figura 10, por lo que la dimensión VC es igual a 3.

FIGURA 10. Separación de 3 vectores por medio de funciones lineales



Fuente: Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular.

Por otro lado, es posible demostrar que si se quieren separar cuatro vectores mediante un conjunto de funciones lineales, estos no pueden ser separados bajo todas las posibles combinaciones:

$$2^4 = \sum_{n=0}^4 \binom{4}{n} = 16$$

Por lo que la dimensión VC de este conjunto de vectores es igual a 3.

APÉNDICE B. FORMACIÓN DE UNA MÁQUINA DE APRENDIZAJE EN TRES DIMENSIONES

Un hiperplano es la generalización del concepto de recta y plano en un espacio  $M$ -dimensional. La ecuación de una recta se define como:

$$y = wx + b$$

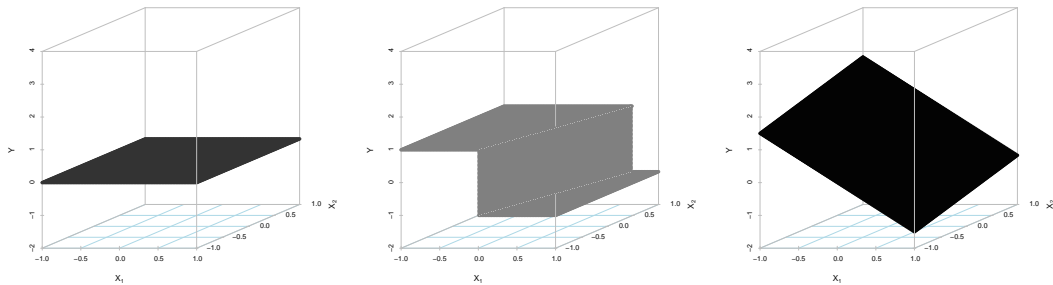
Donde  $w$  es la pendiente de la recta y  $b$  es el punto de corte de la recta con el eje  $y$ . Por otro lado la definición de un plano esta dada por:

$$y = w_1x_1 + w_2x_2 + b$$

Teniendo en cuenta que el hiperplano de un espacio de tres dimensiones es un plano y el de dos es una recta, entonces un hiperplano se puede considerar como un separador. Por ejemplo, en la Figura 11 la primera gráfica muestra un plano que separa un cubo.<sup>20</sup> Por lo tanto, un hiperplano de un espacio de dimensión  $M$  es un separador de dimensión  $M - 1$  que divide el espacio en dos.

A continuación se presenta la construcción de un hiperplano de separación para un vector de características de dos dimensiones  $(x_1, x_2)$  con  $-1 \leq x_1, x_2 \leq 1$ . En la Figura 11 se presentan gráficas de las funciones que se utilizan para construir dicho hiperplano donde  $w_1 = -1,5$ ,  $w_2 = 0,5$  y  $b = 0,5$ : La gráfica de la izquierda de la Figura 11 muestra un plano que separa el espacio, este plano esta

FIGURA 11. Representación en 3D de la clasificación a partir de un hiperplano



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. Para la construcción de las gráficas se tomaron los valores  $w_1 = -1,5$ ,  $w_2 = 0,5$  y  $b = 0,5$ .

definido por:

$$w_1x_1 + w_2x_2 + b = 0$$

La gráfica del centro representa una función indicadora análoga a (4):

$$\begin{aligned} w_1x_1 + w_2x_2 + b &\geq +1 \text{ para } y_k = +1 \\ w_1x_1 + w_2x_2 + b &\leq -1 \text{ para } y_k = -1 \end{aligned}$$

<sup>20</sup>En el caso de una dimensión el hiperplano es un punto que divide una recta en dos y en el caso de dos dimensiones el hiperplano es una recta que cumple la misma función.

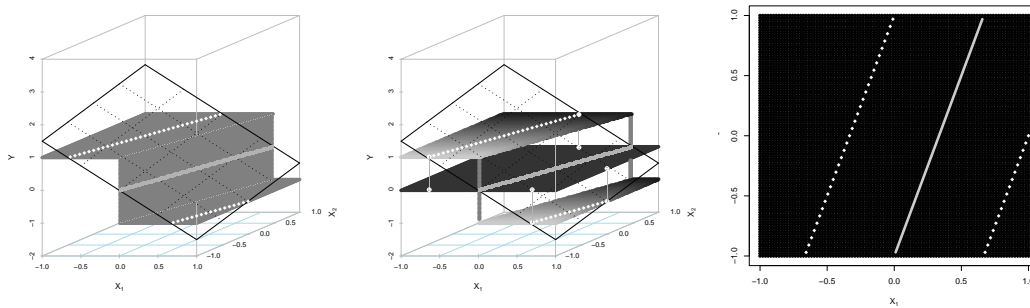
La gráfica de la derecha, denominada función de decisión, esta dada por la siguiente expresión:

$$w_1x_1 + w_2x_2 + b = y$$

La interacción entre estas funciones permite llevar a dos dimensiones el desarrollo de SVM descrito en la sección 4.

La Figura 12 muestra el proceso de formación de los vectores de soporte para el vector asociado a las características  $(x_1, x_2)$ .

FIGURA 12. Representación en 3D de la clasificación a partir de un hiperplano



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular. Para la construcción de las gráficas se tomaron los valores como dados;  $w_1 = -1,5$ ,  $w_2 = 0,5$  y  $b = 0,5$ .

La gráfica de la izquierda de la Figura 12 muestra la intersección entre la función indicadora y la función de decisión. Esta define el margen de separación que está dado por los posibles vectores de soporte (puntos blancos). A su vez, este conjunto genera el hiperplano de separación (línea gris gruesa) que a su vez hace parte de la intersección con el plano que divide el cubo.

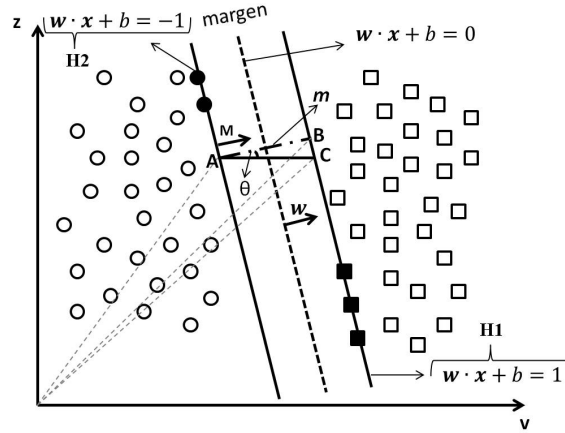
La gráfica del centro presenta la proyección de las rectas formadas por los vectores de soporte y que generan los hiperplanos de separación en dos dimensiones (plano negro); es decir, de las líneas que separan el plano y forman el margen de separación de la máquina de aprendizaje.

Por último, la gráfica de la derecha presenta el resultado final donde se tienen los hiperplanos que caracterizan la SVM. Hay que resaltar que los valores de  $w_1$ ,  $w_2$  y  $b$  deben ser estimados a partir de los datos utilizando las técnicas discutidas en la sección 4.

#### APÉNDICE C. MARGEN DE SEPARACIÓN

Sea  $A$  un punto de coordenadas  $\{v_1, z_1\}$  que hace parte del hiperplano  $H_2$ ,  $B$  y  $C$  puntos de coordenadas  $\{v_2, z_2\}$  y  $\{v_3, z_3\}$ , respectivamente, que hacen parte de  $H_1$  tal como se presenta en la Figura 13; entonces  $A$ ,  $B$  y  $C$  forman el triángulo rectángulo  $\triangle ABC$  cuyo ángulo  $\widehat{BAC}$  es  $\theta$ .

FIGURA 13. Problema de clasificación a partir de un hiperplano óptimo



**Fuente:** Gráfica ilustrativa realizada por los autores. Esta no representa algún conjunto de datos en particular.

Además, sea  $M$  tal que  $\|M\| = m$  y  $M = \alpha w$  de forma que  $M$  es paralela a  $w$ , entonces mediante propiedades trigonométricas se tiene que:

$$m = \cos(\theta) \|C - A\|$$

Dado que  $\cos(\theta) = \frac{M \cdot (C - A)}{\|M\| \cdot \|C - A\|}$ , se tiene que:

$$\begin{aligned} m &= \frac{M \cdot (C - A)}{\|M\| \cdot \|C - A\|} \|C - A\| \\ m &= \frac{\alpha w}{\|\alpha w\|} (C - A) \\ m &= \frac{w}{\|w\|} (C - A) \end{aligned} \quad (28)$$

Entonces, se tiene que la resta entre de las ecuaciones (2) y (3) evaluadas en las cotas que definen la igualdad para dos puntos  $A$  y  $C$  (Figura 13) es igual a:

$$w(C - A) = 2$$

Finalmente, reemplazando en (28) se tiene que:

$$m = \frac{2}{\|w\|}$$

APÉNDICE D. OTRAS FUNCIONES *Kernel*

A continuación se presentan algunas funciones *kernel* más utilizadas según Ivanciuc (2007). Estas dependen de algunos parámetros que se encuentran mediante diferentes métodos empíricos o estadísticos:

- **Exponential Radial Basis Function Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma^2}\right)$$

- **Neural (Sigmoid, Tanh) Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a\mathbf{x}_i \cdot \mathbf{x}_j + b)$$

- **Anova Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_i \exp(-\gamma(\mathbf{x}_i - \mathbf{x}_j))\right)^d$$

- **Fourier Series Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \frac{\text{sen}(N + 1/2)(\mathbf{x}_i - \mathbf{x}_j)}{\text{sen}(1/2(\mathbf{x}_i - \mathbf{x}_j))}$$

- **Spline Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{r=0}^k \mathbf{x}_i^r \mathbf{x}_j^r + \sum_{s=1}^N (\mathbf{x}_i - \mathbf{t}_s)_+^k (\mathbf{x}_j - \mathbf{t}_s)_+^k$$

- **Additive Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_n K_n(\mathbf{x}_i, \mathbf{x}_j)$$

- **Tensor Product Kernel:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \prod_n K_n(\mathbf{x}_i, \mathbf{x}_j)$$

## APÉNDICE E. DESCRIPCIÓN DE LOS DATOS

Los datos de Frank y Asuncion (2010) contienen la información crediticia de 1000 clientes para un banco alemán. La base contiene 20 variables que describen características propias de cada cliente y de las condiciones iniciales del crédito. Las variables son cualitativas y numéricas. A continuación se presenta una breve descripción de cada una de las variables utilizadas dentro del análisis<sup>21</sup>:

**Estado de la cuenta:** Indica el estado del crédito de un cliente del banco, puede ser cumplimiento o incumplimiento.

**Duración:** Duración en meses del crédito.

**Historia crediticia:** Corresponde a variables que miden la historia crediticia del cliente, es decir, el comportamiento de los clientes con los créditos actuales y pasados.

**Propósito del crédito:** Corresponde al destino del crédito. Por ejemplo, vehículo o vivienda.

**Monto del crédito:** Corresponde al monto solicitado en el crédito.

<sup>21</sup>Una descripción detallada de los datos se encuentra en [http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

**Cuentas de ahorro:** Corresponde al monto que se encuentra en las cuentas de ahorro de los clientes.

**Duración en el empleo actual:** Tiempo en el empleo actual del cliente. Se clasifica en rangos.

**Interés respecto al ingreso disponible:** Razón entre el monto del interés del crédito y el ingreso disponible de los clientes.

**Estado civil y género:** Esta variable reúne el estado civil del cliente y su género.

**Otros deudores o garantías:** Indica si existen otros deudores y garantías del crédito.

**Tiempo de residencia en hogar actual:** Tiempo de residencia del cliente en su hogar actual.

**Edad:** Edad en años.

**Propiedades:** Tipo de propiedades con las que cuenta el cliente.

**Otras deudas:** Indica si el cliente presenta otras deudas financieras como el pago de créditos a otras entidades.

**Tipo de vivienda:** Corresponde al tipo de vivienda del cliente, propia o en alquiler.

**Número de créditos con el banco:** Corresponde al número de créditos que ha tenido el cliente con el banco.

**Tipo de trabajo:** Es una variable que clasifica al cliente según su trabajo.

**Personas que dependen del cliente:** Número de personas que dependen económicamente del cliente.

**Celular:** Indica si el cliente tiene o no un celular.

**Trabajador extranjero:** Indica si el cliente es extranjero.

Los datos para el caso colombiano corresponden al formato 341 de la Superintendencia Financiera de Colombia. Estos fueron proporcionados por los Departamentos de Estabilidad Financiera y de Información Técnica y Económica del Banco de la República. El formato 341 registra información de endeudamientos de acuerdo a las normas de evaluación de operaciones activas de créditos. Bajo lo anterior, las variables que se tomaron para el análisis reflejan características propias del crédito más que del cliente en sí. Para nuestro caso se tomaron créditos comerciales y aquellas variables que se consideran relevantes para el análisis. Las variables se describen brevemente a continuación<sup>22</sup>:

**Estado de la cuenta:** Indica el estado del crédito. Si los días de mora son mayores a 90 entonces se dice que un cliente incumple, de lo contrario no.

**Número de operaciones con el banco:** Corresponde al número de operaciones que se tienen actualmente con el banco.

**Tipo de entidad prestadora:** Corresponde al tipo de entidad que otorga el préstamo. Pueden ser establecimientos bancarios, corporaciones financieras, compañías de financiamiento comercial, banca de segundo nivel, entre otros.

**Tipo de entidad deudora:** Corresponde al tipo de entidad a la que se le otorga el préstamo. Por ejemplo si es pública o privada y a qué tipo de sociedad corresponde.

**Tipo de garantía:** Corresponde al tipo de garantía del crédito, si es idónea o no.

**Capital:** Corresponde al monto del capital de la deuda.

**Intereses:** Corresponde al monto de los intereses de la deuda.

**Garantía:** Corresponde al monto de la garantía.

**Plazo:** Corresponde al plazo de la deuda en meses.

<sup>22</sup>Una descripción detallada del formato 341 y el contenido del mismo se encuentra en: [http://www.superfinanciera.gov.co/NormativaFinanciera/Archivos/ance048\\_07.zip](http://www.superfinanciera.gov.co/NormativaFinanciera/Archivos/ance048_07.zip)



**Peso de los intereses:** Corresponde a la razón entre el monto de los intereses y el capital de la deuda.

A continuación se indica el tipo de variables que componen las dos bases de datos analizadas:

CUADRO 4. Variables bases de datos

Base de Datos	Variable	Tipo
Alemania - Individuos	Estado de la deuda	Cualitativa
	Duración	Numérica
	Historia crediticia	Cualitativa
	Propósito del crédito	Cualitativa
	Monto del crédito	Numérica
	Cuentas de ahorro	Cualitativa
	Duración en el empleo actual	Cualitativa
	Intereses respecto al ingreso disponible	Numérica
	Estado civil y género	Cualitativa
	Otros deudores o garantías	Cualitativa
	Tiempo de residencia en hogar actual	Numérica
	Edad	Numérica
	Propiedades	Cualitativa
	Otras deudas	Cualitativa
	Tipo de vivienda	Cualitativa
	Número de créditos con el banco	Numérica
	Tipo de empleo	Cualitativa
	Personas que dependen del cliente	Numérica
	Celular	Cualitativa
Trabajador extranjero	Cualitativa	
Colombia - Empresas	Estado de la deuda	Cualitativa
	Número de operaciones con el banco	Numérica
	Tipo de entidad prestadora	Cualitativa
	Tipo de entidad deudora	Cualitativa
	Tipo de garantía	Cualitativa
	Capital	Numérica
	Intereses	Numérica
	Garantía	Numérica
	Plazo	Numérica
Peso de los intereses	Numérica	