SVM para la clasificación de asegurados en el seguro del automóvil

Piedad Tolmos Rodríguez-Piñero, piedad.tolmos@urjc.es, Universidad Rey Juan Carlos

RESUMEN

La clasificación de los asegurados en grupos homogéneos de riesgo de cara al cálculo de una prima justa, es parte del proceso de tarificación en seguros de cualquier ámbito. Las técnicas que se emplean habitualmente son de tipo estadístico. En este artículo se presentará un método novedoso para ese propósito, las Máquinas de Vectores Soporte (SVM). El objetivo será clasificar a los clientes de una conocida aseguradora del seguro del automóvil[1] en dos grupos, atendiendo a si presentan o no siniestro en un periodo de un año. Los asegurados vendrán descritos por los valores de los factores de riesgo seleccionados por la aseguradora. Los resultados obtenidos se compararán con los que se consiguen tras la aplicación de un análisis discriminante.

PALABRAS CLAVE: SVM, Clasificación, Factores de riesgo, Análisis Discriminante.

ABSTRACT:

In this paper we propose an approach to classify an insured of a given automobile portfolio according to report a claim or not during a year. This task is developed usually making use of statistical techniques. The algorithm we suggest is based on the application of a standard Support Vector Machines (SVMs), used to classify insures as failed (reported a claim) or not failed. We used with this purpose given risk factors included in the data base of the insurance company MAPFRE. In the simulations section, we compare the performance of the SVM and of a classical statistical method, Discriminant Analysis. The results obtained shown that the learning techniques, and SVM in particular, can be useful tools for researchers interested in evaluating claim frequency of insureds of automobile insurance firms.

KEYWORDS: classification, risk factors, support vector machines, discriminant analysis.

1. INTRODUCCIÓN

Cuando el número de pólizas en una compañía aseguradora es lo suficientemente grande, el desarrollo de un sistema de clasificación adecuado es el primer paso para lograr una prima justa. Se trata de clasificar a sus clientes del modo más homogéneo posible atendiendo al riesgo, de manera que los asegurados pertenecientes a un mismo grupo paguen idéntica prima.

---

[1] La base de datos utilizada en el estudio ha sido cedida por la Fundación MAPFRE Estudios en el ámbito de la Beca de Riesgos y Seguros 2006 que nos fue concedida.

En la literatura, se pueden encontrar soluciones a esta tarea empleando métodos estadísticos. Las técnicas del análisis estadístico multivariante son las que permiten organizar procesos de selección teniendo en cuenta simultáneamente el conjunto de factores de riesgo.

No hemos encontrado apenas referencias del uso de las estrategias de *aprendizaje* en el campo actuarial. Técnicas tales como las Redes Neuronales Artificiales, o las Máquinas de Vectores Soporte han demostrado ser unos clasificadores excelentes en términos de la llamada *tasa de clasificación*, y se han aplicado con éxito en multitud de problemas complejos, como el del reconocimiento de caracteres escritos, la "limpieza" de imágenes, minería de datos, diagnósticos médicos, etc. Sin embargo, en el campo de los seguros, su uso se ha centrado casi exclusivamente en el estudio de la predicción de insolvencia en empresas de seguros y temas relacionados [Shapiro 2001], [Shapiro 2002], [Salcedo-Sanz et al. et al. 2005]. Es por ello, por la aplicación novedosa que suponen, y la confianza en obtener buenos resultados, que nos ha parecido interesante el emplearlas en la resolución de nuestro problema de clasificación.

La cuestión que nos planteamos resolver es la siguiente: dada una cartera de asegurados de una conocida empresa del seguro del automóvil, descritos por sus factores de riesgo, pretendemos clasificarlos en dos clases, atendiendo a si han presentado o no siniestros en el periodo de un año. De este modo, dado un asegurado nuevo, seremos capaces de asignarlo a una de esas dos clases, o lo que es lo mismo, de predecir si va a tener o no un siniestro en ese año.

El trabajo comenzará por formular matemáticamente el problema, describiendo seguidamente las Máquinas de Vectores Soporte. Se pasará entonces a la ejecución de los experimentos, comentando los resultados alcanzados. Los resultados de la clasificación se compararán con los obtenidos aplicando una técnica estadística válida para abordar este tipo de cuestiones, el Análisis Discriminante. Se finalizará con las conclusiones que se extraen de lo presentado en el artículo, y con el resumen de algunas de las aplicaciones que estamos realizando en la actualidad.

2. FORMULACIÓN DEL PROBLEMA

El tipo de problema que planteamos se puede englobar dentro de lo que se conoce como *problemas de clasificación con múltiples atributos* cuya tarea básica es asignar un *objeto*, descrito por los valores que toman ciertos *atributos*, a una serie de *clases*.

Matemáticamente, la representación de un problema de clasificación con múltiples atributos es la siguiente [Schölkopf 1999]: se quiere estimar una función (de decisión) $f : \Re^n \to \{\pm 1\}$ empleando datos (observaciones u *objetos*) del conjunto de observaciones que se utilizará para *entrenar* lo que se conoce como "máquina de clasificación" (red neuronal, algoritmo genético,…). Consideraremos para ello una serie de objetos $\{\mathbf{x}_i\}, \mathbf{x}_i \in \Re^n$, $i \in \{1,...l\}$ generados por cierta función de distribución de probabilidad desconocida $P(\mathbf{x},y)$, donde $y_i \in \{1,-1\}$ constituyen el conjunto de "etiquetas" asociadas (es la salida, la que indica a qué clase pertenece cada $\mathbf{x}_i$). De este modo, el conjunto de datos considerados sería

$$\left(\mathbf{x}_1, y_1\right),\ldots,\left(\mathbf{x}_l, y_l\right) \in \Re^n \times \{\pm 1\} \qquad 2.1$$

y el objetivo es que la función *f* clasifique correctamente los ejemplos nuevos que se la presenten (**x**,y), esto es, que *f*(**x**)=y para ejemplos (**x**,y) generados por la misma distribución de probabilidad *"subyacente"* *P*(**x**,y) que los datos utilizados para el entrenamiento.

El *error de entrenamiento* se puede definir como:

$$R_{ent}\left[\alpha\right] = \frac{1}{l}\sum_{i=1}^{l}\frac{1}{2}\left| f(x_i,\alpha) - y_i \right| \qquad 2.2$$

El error de test (*riesgo*) esperado para una máquina de entrenamiento se define como

$$R[\alpha] = \int \frac{1}{2}\left| f(x,\alpha) - y\right| dP(x,y) \qquad 2.3$$

La cantidad $\frac{1}{2}\left| f(x_i,\alpha) - y_i \right|$ recibe el nombre de *pérdida*.

2.1 EL PROBLEMA DE LA CLASIFICACIÓN DE LOS ASEGURADOS

Para predecir la siniestralidad de un asegurado, vamos a separar a todos los clientes en dos clases: la de los que tendrán siniestros, y la de los que no. Por ello, lo trataremos como un problema de clasificación con múltiples atributos "simple", esto es, con sólo dos clases. Los conjuntos $\{\mathbf{x}_i\},\ i \in \{1,\ldots l\}\ \mathbf{x}_i \in \Re^n$ representan a los asegurados, descritos por un conjunto de *n* factores de riesgo (cada componente de $\mathbf{x}_i$, $\mathbf{x}_{ij}$, es un factor), y las etiquetas $y_i \in \{-1,1\}$ indicarían la clase, -1 si no presentan siniestros, y 1 en caso contrario. De este modo, durante el periodo de entrenamiento estaríamos manejando pares del tipo $(\mathbf{x_i}, y_i)$, con y_i conocida; el objetivo será, recordemos, que se clasifiquen correctamente los ejemplos nuevos que se presenten $(\mathbf{x}, y)$ generados por la misma distribución de probabilidad *"subyacente"* $P(\mathbf{x}, y)$ que los datos utilizados para el entrenamiento del clasificador

2.2 LOS DATOS UTILIZADOS

Para la ejecución se ha empleado una muestra de 58238 asegurados obtenida al enlazar la cartera de Clientes con la de Siniestros del año 2003 que nos proporcionó MAPFRE.

Los factores de riesgo que describían a cada asegurado eran originalmente 13: Antigüedad carnet, Edad, Tipo carnet, Sexo, Estado Civil, Profesión, Antigüedad vehículo, Uso, Zona Circulación, Potencia, Valor

Sin embargo, hubo que segregarlos para obtener variables categóricas, como exigía el sistema, manejando finalmente 105 variables de entrada. La salida, recuérdese, sólo tomaba dos valores, 1 o -1.

3. LAS MÁQUINAS DE VECTORES SOPORTE

La SVM es una nueva técnica de clasificación, que ha demostrado sobradamente su capacidad de resolución frente a problemas de elevado grado de complejidad. Diseñadas en principio para tratar problemas de clasificación binarios (en dos grupos), se trata de una máquina de aprendizaje que implementa la siguiente idea: cuando no sea posible separar los datos en el espacio de entrada con un hiperplano lineal, trasladar, mediante una aplicación no lineal, los vectores de entrada a un nuevo espacio de dimensión muy alta. En este nuevo espacio se construirá una superficie de decisión lineal. Las especiales propiedades que poseerá esta superficie garantizarán que la capacidad de generalización de la máquina de aprendizaje sea alta. Aunque esta idea se empleó en los primeros experimentos para datos que podían separase sin errores, se puede extender para el caso no separable con notable éxito. La parte conceptual del problema la resolvió Vapnik en 1965 para el caso de *hiperplanos óptimos* para clases separables. En este contexto, Vapnik definió un hiperplano óptimo como una función de decisión lineal con el margen de separación máximo entre los vectores de las dos clases. Se observó entonces que para construir tal hiperplano, uno sólo debía tener en cuenta una cantidad pequeña de los datos de entrenamiento, los llamados *vectores soporte*, quienes determinaban ese *margen*.

Sea un conjunto de asegurados representado por sus factores de riesgo expresados mediante los vectores $\{\mathbf{x}_i\}, i \in \{1,...l\}$, y un conjunto de etiquetas asociadas $y_i \in \{-1,1\}$[2] que determinan a qué clase pertenece cada asegurado. Supongamos que es posible separar este conjunto de entrenamiento mediante un hiperplano lineal. Los puntos $\mathbf{x}$ que pertenecen al hiperplano satisfacen la ecuación $\mathbf{w} \cdot \mathbf{x} + b = 0$, donde $\mathbf{w}$ es un vector normal al hiperplano, y $|b|/\|\mathbf{w}\|$ es la distancia perpendicular del hiperplano al origen (tomamos $\|\cdot\|$ como la norma euclídea) De entre todos los hiperplanos capaces de separar los datos, existe un único *hiperplano óptimo*, en el sentido de que es capaz de separar los puntos con el mayor margen de separación entre cada elemento del conjunto de entrenamiento y el hiperplano. En este sentido, el algoritmo de aprendizaje diseñado por Vapnik et al., la SVM, resuelve el siguiente problema:

"Encontrar $\mathbf{w} \in \Re^n$ y $b \in \Re$ que

minimicen $\tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$ 3.1

sujeto a $y_i\left(\mathbf{w}^t \cdot \mathbf{x}_i + b\right) \geq 1 \qquad \forall\ i = 1,..., l$" 3.2

Cuando se hallen $\mathbf{w}$ y b, la regla de clasificación para los asegurados será, simplemente, $sign\left(\mathbf{w}^t \cdot \mathbf{x}_i + b\right)$[3], y el error de clasificación cometido vendrá dado por $R_{ent}(\mathbf{w}, b)$, tal y como se describió anteriormente. Por otro lado, aquellos puntos que verifican la igualdad en la inecuación 3.2, y cuya eliminación cambiaría la solución que encontremos, son los que llamaremos ***vectores soporte*** . Estos

---

[2] Describimos el caso más sencillo de dos únicas clases, pues para el general de K clases basta con tomar, como ya se ha visto y_{ik} en vez d y_i.

[3] Esto se corresponde con las funciones de decisión del tipo $f(\mathbf{x}) = sign\left(\mathbf{w}^t \cdot \mathbf{x}_i + b\right)$.

vectores, pertenecerán a uno de los dos posibles hiperplanos óptimos de separación de los que hablábamos antes, representados por las ecuaciones $\mathbf{w}^t \cdot \mathbf{x}_i + b = 1$ para un hiperplano, y $\mathbf{w}^t \cdot \mathbf{x}_i + b = -1$ para el otro.

Si tratamos de aplicar el algoritmo anterior a datos no separables, no encontraremos ninguna solución factible, pues la función objetivo crece desmesuradamente. Para evitarlo, se relaja la restricción 3.2 cuando sea necesario, lo que se logra mediante la introducción de unas variables nuevas de pequeño tamaño. La formulación del problema queda ahora:

“ Encontrar $\mathbf{w} \in \Re^n$ y $b \in \Re$ y ξ_i $i = 1,\ldots,l$ que

minimicen $$\tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \qquad 3.3$$

sujeto a $$y_i\left(\mathbf{w}^t \cdot \mathbf{x}_i + b\right) \geq 1 - \xi_i \quad \text{y} \quad \xi_i > 0 \quad \forall\, i = 1,\ldots, l \text{”} \qquad 3.\ 4$$

donde C es un parámetro que el clasificador deberá estimar.

Por último, en el caso de la SVM no lineal, se proyectan las variables de entrada en un espacio de dimensión mayor (normalmente de dimensión infinita) que aquel al que pertenecían dichas variables, y se aplica la SVM descrita anteriormente en este nuevo espacio, conocido como *espacio de características*. De este modo, la SVM no lineal es capaz de separar los asegurados con una probabilidad de error dada por $R_{ent}(\mathbf{w}, b)$. Esa proyección se realiza utilizando las funciones *núcleo* (kernel).

4. CLASIFICACIÓN DE LOS ASEGURADOS CON UNA SVM

En este experimento utilizamos un software gratuito para la SVM llamado Libsvm[4], con un *kernel* de base radial, $k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\|\mathbf{x} - \mathbf{x}_i\|^2 / c\right)$. Escogimos como método para determinar el error en el conjunto de Test el de la *Validación Cruzada*[5] en 5 *pliegues.*

Los resultados se presentan en forma de la Tasa de Clasificación [(Nº de aciertos) / (Nº de casos)] y de la Matriz de Confusión, en cuya diagonal principal aparecen los casos acertados, y en la secundaria los errores cometidos en cada clase.

77.72%

En términos de la matriz de confusión, el resultado sería:

-1 -> 76.41% (21638/28318)

---

[4] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM : a library for support vector machines*. Software disponible en www.csie.ntu.edu.tw/~cjlin/libsvm

[5] Dividir el conjunto de datos de entrada en n subconjuntos, entrenar el modelo con n-1 de los n conjuntos, y validar los resultados con el conjunto restantes. Repetir el proceso para cada uno de las n posibles elecciones del conjunto omitido

1 -> 79,13%   (23675/29919)

Matriz de Confusión:

21638	7320
6244	23675

Obsérvese como el nº de clasificados correctamente entre los que han tenido siniestro es ligeramente superior al de los que no tuvieron. Esto es lógico, debido a que el nº de asegurados que presentaron siniestros es también un poco mayor que el de los que no lo hicieron.

5. CLASIFICACIÓN DE LOS ASEGURADOS CON ANÁLISIS DISCRIMINANTE. COMPARACIÓN

De entre todas las técnicas utilizadas habitualmente en este tipo de cuestiones, se ha seleccionado el *Análisis Discriminante* para clasificar a los asegurados y comparar así lo obtenido con los que se alcanzaron vía la SVM anterior. De este modo, se podrá comprobar la fiabilidad de los resultados logrados con la nueva herramienta. La elección de este método y no otro, se justifica por su relación directa con las técnicas que hemos denominado de aprendizaje máquina, ya que el Análisis Discriminante, tal y como lo concibió Fisher, se considera un algoritmo de aprendizaje de entre los “estadísticos clásicos”.

Por otra parte, debido a la naturaleza de las variables que se tratan, se debe llevar a cabo un *Análisis Factorial* antes de proceder a discriminar.

Efectivamente, el análisis discriminante es una herramienta que se aplica solamente a variables cuantitativas viniendo caracterizados los grupos a través de una variable categórica, que en nuestro caso será la variable *siniestros*. Las variables del estudio son en su mayor parte de tipo cualitativo. Se hará necesario por tanto, para aplicar un análisis discriminante, transformar previamente el conjunto de variables en variables cuantitativas/continuas.

Para ello se procedió de la siguiente forma:

1.- Realización de un Análisis Factorial Múltiple (AFM) sobre el conjunto completo de las variables.

2.- Selección de aquellos factores resultantes del AFM que expliquen el 100% de la variabilidad total.

3.- Realización del Análisis Discriminante (AD) con los factores anteriormente seleccionados.

El AFM es una técnica de análisis factorial que trata tablas en las cuales un conjunto de individuos viene descrito por varios grupos de variables. En el seno de un mismo grupo, las variables deben ser del mismo tipo (continuo o nominal) pero de un grupo a otro, las variables pueden ser de diferentes tipos. En nuestro experimento se lleva a cabo un AFM sobre el conjunto de individuos descritos a través de dos tablas, la primera formada por las variables continuas y la segunda formada por las variables nominales.

El método es en realidad un análisis factorial del conjunto de grupos (llamado análisis global). Para las variables continuas, el AFM se comporta como un Análisis en Componentes Principales; para las variables nominales como un Análisis de Correspondencias Múltiples. Es la introducción de pesos en las variables, que equilibran las inercias axiales máximas de los grupos, lo que hace posible la presencia simultánea de variables continuas y nominales. El objetivo final es obtener los principales factores de variabilidad de los individuos, estando éstos descritos de forma equilibrada por varios grupos de variables.

Los resultados de aplicar el SPSS se expresan, como ya hicimos antes con la SVM, en una matriz de confusión. Como aquella, el grupo al que de hecho pertenece el asegurado está en las filas, y el pronosticado se recoge en las columnas. La diferencia con la que contenía los resultados de la SVM es que aquí se pueden observar también los porcentajes de clasificados. Concretamente, la tabla ofrece las frecuencias absolutas, los porcentajes de fila y el porcentaje total de clasificaciones correctas, de modo que en la diagonal principal aparece el porcentaje de clasificaciones correctas. Así, la función consigue clasificar correctamente un 69,1% de los casos de los asegurados que no tuvieron siniestro, y un total de 71,8% de los que sí. De este modo, podemos decir que el resultado fue que los 59 factores clasificaban correctamente los siniestros en el 70,5% de los casos.

Resultados de la clasificación[a]

			Grupo de pertenencia pronosticado		
		SINIESTR	-1,00	1,00	Total
Original	Recuento	-1,00	19581	8737	28318
		1,00	8428	21491	29919
	%	-1,00	69,1	30,9	100,0
		1,00	28,2	71,8	100,0

a. Clasificados correctamente el 70,5% de los casos agrupados originales.

5.2 COMPARACIÓN

La siguiente tabla compara los resultados alcanzados con las dos técnicas:

	SVM			AD		
Tasa de clasificación	77.72%.			70,5%		
Matriz de Confusión		-1	1		-1	1
	-1	21638 76,41%	7320	-1	19581 69.1%	8737
	1	6244	23675 79,13%	1	8428	21491 71,8%

Se observa así que la SVM obtiene una tasa de clasificación mayor que el AD. La mejoría es de un 7%, lo que en una Tasa que ya está por encima del 70 %, es realmente importante. Las diferencias entre los elementos acertados de cada clase se aprecian en la matriz de confusión.

6. OTROS RESULTADOS

En lo expuesto en este artículo nos hemos centrado en la clasificación de los asegurados en dos clases. Un problema de clasificación de este estilo lleva parejo un segundo problema, el de la Selección de Características (factores de riesgo, en este caso). Efectivamente, con el objeto de mejorar la capacidad de generalización del clasificador, su estabilidad, y el tiempo de computación, a menudo es necesario, si el nº de variables es grande, realizar una selección de las variables importantes, las que retiene la mayor cantidad de información. Esta cuestión tiene además aquí un valor añadido, en cuanto al interés que la información sobre los factores de riesgo realmente relevantes de cara a la siniestralidad, pueda tener para la aseguradora. Por eso, se han realizado otros experimentos de clasificación seleccionando previamente los factores de riesgo más relevantes. Para ello se han empleado Algoritmos Genéticos y Árboles de Clasificación, comparando los resultados, tanto en cuanto a la clasificación como a la selección.

Más interesantes han resultado las prácticas que se han realizado con una nueva base de datos, correspondiente al año 2005. En este caso, los factores de riesgo que había recogido la aseguradora eran muy diferentes a los que aparecían en la base de 2005, e incluían una variable, el nivel de *Bonus Malus*, que llevó a la ejecución de experimentos levemente diferentes. Se trataba de ver la influencia que tenía a la hora de seleccionar factores, y a la de clasificar. Por ello, seleccionó con un *Random Forest* incluyendo y sin incluir esta variable en la entrada, y una clasificación de los asegurados con el mismo criterio. En la selección, se vio claramente que la influencia del nivel de Bonus Malus era considerable, resultando escogidos muchos de estos niveles. La clasificación, por encima del 70 % en ambos casos, resultaba mejor si se añadía el nivel de Bonus Malus que en caso contrario.

7. CONCLUSIONES. LÍNEAS ABIERTAS DE INVESTIGACIÓN

En el presente artículo se han introducido técnicas novedosas, tomadas del Aprendizaje Máquina, para resolver un problema de Tarificación a priori. Concretamente, se trata de clasificar un grupo de asegurados descritos por sus factores de riesgo en dos clases, atendiendo a si presentan o no siniestro en el periodo de un año.

La clasificación de los asegurados se ejecutó utilizando una SVM, obteniéndose una tasa de clasificación del 77’72%. Este resultado es mejor que el que se logra clasificando mediante un Análisis Discriminante. De este modo, la nueva técnica consigue superar a una de las herramientas estadísticas que se utiliza en este tipo de cuestiones.

En cuanto a futuras investigaciones, este nuevo campo de actuación, el actuarial, ofrece muchas expectativas para la aplicación de las técnicas de aprendizaje, de modo que las líneas de investigación que se abren ante nuestros ojos son numerosas.

En el problema concreto de clasificación de asegurados, quizás la más atractiva sea la de dar el paso a la clasificación multiclase. Si bien no podemos esperar grandes resultados frente a los ya obtenidos, lo natural es que ahora nos planteemos clasificar los asegurados en clases según su riesgo para así poder ser capaces de calcular la prima a pagar en cada una de ellas. Ya se han realizado nuevos experimentos en este sentido con la base de datos de 2005, en la que se ha tenido en cuenta la frecuencia de los siniestros.

Otra cuestión muy interesante, en la que de hecho estamos trabajando actualmente, es una clasificación de los asegurados atendiendo al riesgo de presentar o no siniestro según el factor EDAD. Como ya hemos comentado, este es uno de los factores de riesgo que más interesa a la Compañía aseguradora. Hemos adoptado el método de “uno frente al resto”, de manera que vamos dividiendo las edades por tramos y construyendo clasificadores binarios en cada uno de ellos.

El uso de técnicas de aprendizaje diferentes a las ya aplicadas, como los propios árboles para clasificar, o el *simulated annealing* para seleccionar factores, es otra vía de investigación prometedora.

Además, está la cuestión de interpretar los factores que se obtienen con el Análisis factorial, y poder así compara con los que se consiguen con el Algoritmo Genético y el Árbol de Clasificación que antes se comentaba.

8. BIBLIOGRAFÍA

[Boj 2003] Boj del Val, E. (2003) *Análisis multivariante aplicado a la selección de factores de riesgo en la Tarificación* Tesis Doctoral defendida en el Departament de Matemàtica Econòmica, Financera i Actuarial de la Universitat de Barcelona.

[De Vicente et al 2000]De Vicente y Oliva, M., Manera Bssa, J., Blanco Jiménez, F. J. (2000) *Análisis Multivariante para las Ciencias Sociales*. Ed.: Dykinson. Madrid.

[Duda et al 2001] Duda, R.O., Hart, P.E. Stork, D.G. (2001) *Pattern Classification*. John Wiley & Sons

[Salcedo-Sanz et al 2002] Salcedo-Sanz S.,De Prado-Cumplido M., Pérez-Cruz F., Bousoño-Calzón C. *Feature Selection via Genetic Optimization.* ICANN 2002: 547-552

[Salcedo-Sanz et al. et al. 2005] Salcedo-Sanz, S., Fernández-Villacañas J. L., Segovia-Vargas, M. J. and Bousoño-Calzón, C. 2005. Genetic programming for the prediction of insolvency in non-life insurance companies, Computers & OR 32: 749-765

[Sanchis et al. 2003] Sanchis, A., Gil, J. A., and Heras, A. (2003): "El análisis discriminante en la previsión de la insolvencia en la empresa de seguros no-vida", Revista Española de Financiación y Contabilidad 115.

[Scholkopf 2002] Scholkopf B, Smola A. Learning with kernels. Cambridge, MA: MIT Press; 2002.

[Segovia-Vargas 2003] Segovia-Vargas MJ, Salcedo-Sanz S, Bousoño-Calzón C. Prediction of insolvency in non-life insurance companies using Support Vector Machines and genetic algorithms. In: Proceedings of X SIGEF Congress in Emergent Solutions for the Information and Knowledge Economy, León, Spain, 2003.

[Shapiro 2001] Shapiro, A. Soft Computing Applications in Actuarial Science. ARCH. 2001. http://www.soa.org:80/library/arch/2000-09/arch01v113.pdf

[Shapiro 2002] Shapiro, A. The merging of Neural Networks, fuzzi logic and genetic algorithms. 2002, Insurance: Mathematics and Economics 31 (2002) 115-131

[Vapnik 1995] Vapnik, V., Cortes, C. *Support-Vector Networks.* 1995. Machine Learning, 20, 273-297.

[Weston et al. 2000] Weston, H., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. (2000): *Feature Selection for SVMs*, Advances in NIPS 12, MIT Press, 526-532.