

# MÁQUINAS DE SOPORTE VECTORIAL EN EL RECONOCIMIENTO AUTOMÁTICO DEL HABLA

Juan Carlos Roeder Moreno

yoanitus@yahoo.es

Universidad Nacional de Trujillo

Facultad de Ciencias Físicas y Matemáticas

Escuela Académico-Profesional de Informática

## RESUMEN

El reconocimiento automático del habla es un área de investigación de las ciencias de la computación que está tomando gran auge en los últimos tiempos. Para conseguir esto, diversos investigadores realizaron o utilizan diversas técnicas, especialmente las de las máquinas de aprendizaje, dado que con estas técnicas se la máquina podrá discernir o tomar decisiones de acuerdo al aprendizaje. Existen diversas máquinas, entre las cuales están las Máquinas de Soporte Vectorial, consideradas como la vanguardia de las máquinas de aprendizaje, gracias a su poder computacional por poder trabajar con funciones complejas. El presente informe muestra funcionan las SVM y como estas pueden aportar gran ayuda al reconocimiento automático del habla (RAH).

**Palabras Claves:** SVM, Máquinas de Soporte Vectorial, RAH, Reconocimiento Automático del Habla, Red Neuronal, Máquinas de Aprendizaje.

## 1. INTRODUCCION

En la actualidad ha tomado gran auge, la Inteligencia Artificial, para dar solución a problemas diversos. Uno de estos problemas es el del Reconocimiento Automático del Habla, es decir, el de poder reconocer las palabras habladas.[1]

La inteligencia artificial utiliza diversas técnicas para esto, entre las cuales está el Aprendizaje de Máquinas.

El aprendizaje puede ser descrito como el arte de encontrar una regla general que explique una serie de estímulos o eventos, es decir, un conjunto de información tomada del medio.[2]

Existen diversos tipos de máquinas de aprendizaje, de las cuales, las que más destacan son las Máquinas de Aprendizaje Supervisadas y las No Supervisadas.

En una máquina Supervisada, a partir de un conjunto de muestras de entrenamiento, que comprende de datos muestras  $X$  ingresados, más un conjunto de de valores

salida o respuesta  $Y$ , que nos indican las clases en las que pueden ser identificadas los objetos ingresados.

Las máquinas No Supervisadas, a diferencia de las mencionadas anteriormente, estas no tienen valores de salida esperados en el entrenamiento. Estas a partir de los datos ingresados encuentran características comunes, agrupando o separando los objetos o datos de acuerdo a dichas características en diferentes grupos o clases. Esto puede tomarse como el caso de la clusterización.

Entrando en las máquinas supervisadas, existen diversas de ellas. Pero destacan entre ellas las famosas Redes Neuronales y las Máquinas de Soporte Vectorial.

Una red neuronal artificial es un procesador distribuido en paralelo de forma masiva que tiene una tendencia natural para almacenar conocimiento de forma experimental y lo hace disponible para su uso.

Se parece al cerebro humano en dos aspectos:

- ✓ El conocimiento es adquirido por la red a través de un proceso de aprendizaje.
- ✓ Los pesos sinápticos o fuerza con que están interconectadas las neuronas se utilizan para almacenar la información.

Relacionando el campo del reconocimiento de patrones, con las redes neuronales, se pueden crear sistemas los cuales se encarguen de la identificación y reconocimiento de objetos, mediante el aprendizaje.[2]

Las máquinas de Soporte Vectorial, a diferencia de las redes neuronales, abstraen el problema desde un espacio de atributos a un espacio de patrones de características, con mayor dimensión para ser separadas por un hiperplano.

Son consideradas un efectivo clasificador discriminante capaz de minimizar el margen de error, ya que pueden lidiar con ejemplares con una dimensionalidad muy alta. También, maximiza el margen de separación entre clases mediante el hiperplano que las separa, teniendo a los puntos más cercanos, equidistantes a dicho hiperplano.

El presente trabajo trata de explicar el funcionamiento de las SVM y cómo estas pueden ofrecer un buen desenvolvimiento en el RAH.

## 2. MÁQUINAS DE SOPORTE VECTORIAL

Fueron creadas por *Vladimir Vapnik*, que durante los 60's y 90's, desarrolló junto con *Alexey Chervonenkis* la teoría VC, la cual es base fundamental de la teoría de aprendizaje, dado que proporciona una forma para minimizar el riesgo empírico, riesgo que se presenta en toda máquina clasificadora.[3]

Comenzaremos describiendo el caso de las Máquinas de Soporte Vectorial Lineales. [4]

Sea el par de entrenamiento:

$$\{x_i, y_i\}; i = 1, \dots, l; y_i \in \{-1, 1\}; x_i \in R^d$$

Se tiene un hiperplano que divide las clases  $Y$ , de manera que, la ecuación de dicho hiperplano pertenece a  $R^{d-1}$

Los puntos de  $X$  que están encima del hiperplano satisfacen la ecuación

$$w \cdot x + b = 0$$

Donde  $w$  es normal al hiperplano.

Además, se tiene  $\frac{|b|}{\|w\|}$  que es la distancia del hiperplano al origen, y  $\|w\|$  es la norma euclidiana de  $w$ .

Sean  $d_+$  y  $d_-$  la distancia del hiperplano separador al punto más cercano perteneciente a la clase positiva ( $y = 1$ ) y a la clase negativa ( $y = -1$ ) respectivamente.

Se le llamará margen de un hiperplano de separación a la suma  $d_+$  y  $d_-$

Para los casos linealmente separables, el algoritmo de soporte vectorial busca el hiperplano separador con el mayor margen.

Por lo tanto, puede ser expresado de la siguiente manera, bajo el supuesto que los datos de entrenamiento satisfacen las siguientes restricciones:

$$x_i \cdot w + b \geq +1 \quad \text{para } y_i = +1 \dots \dots \dots (1)$$

$$x_i \cdot w + b \leq -1 \quad \text{para } y_i = -1 \dots \dots \dots (2)$$

De las cuales se puede despejar y expresar en una sola inequación:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \dots \dots \dots (3)$$

Considerando que los puntos que satisfacen la ecuación (1) se mantienen (de tal manera, que exista un punto que coincida con un escalamiento a  $w$  y  $b$ ).

Estos puntos pertenecen al hiperplano  $H_1: x_i \cdot w = 1$  con normal  $w$  y distancia al origen  $\frac{|1-b|}{\|w\|}$ . Similarmente, los

puntos para la ecuación (2) se mantienen sobre el hiperplano  $H_2: x_i \cdot w = -1$ , también con normal  $w$ , y distancia  $\frac{|-1-b|}{\|w\|}$ .

Dado que  $d_+ = d_- = \frac{1}{\|w\|}$  y el margen total es  $d_+ + d_- = \frac{2}{\|w\|}$

Observar que  $H_1$  y  $H_2$  son paralelos (tienen la misma normal) y no existen puntos de entrenamiento entre ambas.

Por lo tanto, se puede encontrar el par de hiperplanos que tienen el máximo margen, mediante la minimización de  $\|w\|^2$ , sujeto a las restricciones de (3).

Por lo tanto, se espera la solución de un típico caso bi-dimensional para obtener la forma mostrada en la Figura 01. Estos puntos de entrenamiento los cuales satisfacen la restricción en (3) (es decir, que están por encima de uno de los hiperplanos  $H_1$  y  $H_2$ ) y para los cuales, si son removidos, la solución encontrada cambiaría, son los vectores de soporte.

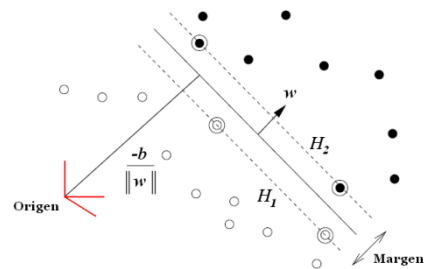


Figura 01. Hiperplanos de separación lineal para el caso separable. Los vectores soportes están rodeados por una circunferencia.

## 3. MÁQUINAS DE SOPORTE VECTORIAL NO LINEALES Y EL KERNEL

Como se vio en el unto anterior, el caso lineal de las SVM consiste en encontrar la máxima separación de los vectores soporte al hiperplano.

Pero esto no siempre sucede. Vapnik demostró que estando en un cierto espacio de dimensión, solo es posible llegar a separar tantos datos como la dimensión más una unidad. Esta demostración la puede encontrar en [3]

Con esto se puede concluir entonces que cierto datos no pueden llegar a ser clasificados en forma lineal.

Vapnik llegó a una solución para este problema. La solución es llevar los datos de una dimensión hacia otra dimensión, haciendo uso de Kernels.

Suponga la función lineal:

$$f(X) = \sum_{j=1}^n w_j \cdot x_j \dots \dots \dots (4)$$

Pero además, suponga que  $W$  se escribe de la forma:

$$W = \sum_{i=1}^m \alpha_i \phi(x_i) \dots\dots\dots (5)$$

Donde la función  $\phi$  es conocida como la función de mapeo de los vectores de entrenamiento.

Las ecuaciones (4) y (5) muestran formas básicas en la clasificación lineal. Escribimos la forma dual reemplazando la ecuación (5) en la ecuación (4)

$$f(X) = \sum_{j=1}^n \left( \sum_{i=1}^m \alpha_i \cdot \phi(x_i) \right) x_j$$

De lo anterior, podemos escribir la función en términos de la combinación lineal de la función de mapeo. Por lo tanto:

$$f(X) = \sum_{i=1}^m \alpha_i \sum_{j=1}^n \phi(x_i) \cdot \phi(x_j) = \sum_{i=1}^m \alpha_i K(x_i, x)$$

Donde  $K$  es llamado Kernel, es considerado como la función Kernel Clasificador y  $\alpha \in R^m$  son coeficientes de expansión.

Lo anterior nos muestra que haciendo uso de una función kernel clasificadora podemos llegar a separar o clasificar los objetos o vectores en un cierto espacio, ya que permite simplificar la función separadora en una forma más simple.

Vapnik usó esta idea para llevar los datos que se encuentran en una dimensión que no pueden ser clasificados en forma lineal, a un estado de dimensión diferente, donde puedan ser clasificados, haciendo uso del método de clasificación lineal.

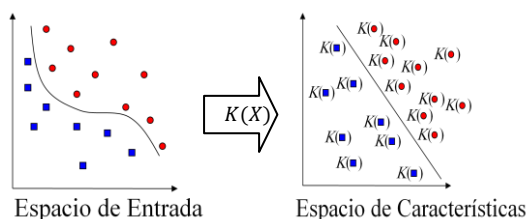


Figura 02. Utilización de una función Kernel para transformar datos para poder ser clasificados en forma lineal.

La idea es transformar los datos que se encuentran en una dimensión, en datos que se encuentran en una dimensión mayor.

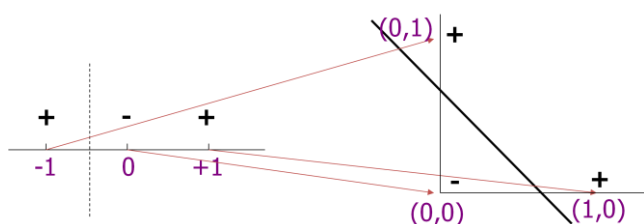


Figura 03. Transformación de puntos de una dimensión  $n$  a una dimensión  $n'$  de mayor dimensionalidad.

Esto se puede interpretar como llevarlos de un espacio donde se encuentran los patrones de entrada, a un espacio de donde se utilizará las características para realizar la clasificación (Figura 02).

#### 4. SVM Y SU APLICACIÓN EN RAH.

En la actualidad, las SVM muestran ser una poderosa forma de clasificación, dado que pueden trabajar con funciones complejas en espacios de alta dimensionalidad.

Esto puede ser aplicado en diferentes campos. En el campo del reconocimiento automático del habla a dado grandes aportes, al mejorar las tasas de reconocimiento. Aunque actualmente las SVM son utilizadas mayormente en híbridos. Esto se muestra en diferentes investigaciones realizadas, como lo muestran por ejemplo *Curban* y *Thiran* [5], al utilizar un híbrido entre los modelos ocultos de Markov y las SVM. También está la investigación de *Wan* y *Carmichael* [6], donde muestran un híbrido de las SVM con el DTW, al utilizar este último como un kernel para llegar a separar linealmente la data.

Estas y otras investigaciones muestran que las SVM llegan a ser muy prometedoras y proporcionan una performance que dan solución al problema del RAH en forma robusta.

En futuras investigaciones se llegará a implementar una SVM para la clasificación de palabras.

#### 5. CONCLUSIONES

- ✓ Las SVM son un concepto relativamente nuevo. Aparecieron en la década de los 90's.
- ✓ Las SVM o Máquinas de Soporte Vectorial, basadas en la teoría de Vapnik, son un tipo de máquinas de aprendizaje, que implementan el principio inductivo de minimización de riesgo estructural de la teoría de aprendizaje estadístico, que en la actualidad tienen una gran aplicación para problemas de regresión, clasificación y estimación de problemas entre otros.
- ✓ En la actualidad la mayoría de aplicaciones de las SVM se dan en la clasificación de patrones para el reconocimiento de objetos, influyendo más en el área del RAH (Reconocimiento Automático del Habla).
- ✓ Las SVM abstraen el problema desde un espacio de atributos a un espacio de patrones de características, con mayor dimensión para ser separadas por un hiperplano.
- ✓ Son capaces de trabajar en dimensiones altas y pueden usar funciones de clasificación complejas.
- ✓ Se encuentran en la vanguardia del campo del aprendizaje de máquinas.

## 6. REFERENCIAS

- [1] <http://elies.rediris.es/elies12/cap241.htm>. Comunidad Virtual de Usuarios asociada la lista de distribución Infoling. 2001.
- [2] *R. Herbrich*, 2002. **Learning Kernel Classifiers, Theory and Algorithms**. The MIT Press, Cambridge, Massachusetts; Londres, Inglaterra, 2002, cap. 1, pp. 1 – 13.
- [3] *V. Vapnik*, 2000. **The Nature of Statistical Learning Theory**. Springer. Edición 2.
- [4] *C. Burges*, 1998. **A Tutorial on Support Vector Machines for Pattern Recognition**. **Data Mining and Knowledge Discovery**. (Junio de 1998). Edición 2, cap. 2, pp. 121 – 167.
- [5] *Gurban M. and Thiran J.* 2005. **Audio-visual speech recognition with a hybrid SVM-HMM system**. 13º Conferencia Europea de Procesamiento de Señales.
- [6] *Wan V., Carmichael J.* 2005. **Polynomial Dynamic Time Warping Kernel Support Vector Machines for Dysarthric Speech Recognition with Sparse Training Data**. Paper. Departamento de Ciencias de la Computación, Universidad de Sheffield, UK.
- [7] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> Chih-Chung Chang y Chih-Jen Lin. Concejo Nacional de Ciencia de Taiwan.