Máquinas de Vectores de Soporte

M.Sc. Angelo Jonathan Diaz Soto

(SVM)

(Conceptos básicos)

2025

[redacted]

Contenido

1 [redacted]

2 [redacted]

(Data &

3 [redacted]

4 [redacted]

5 [redacted]

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Support Vector Regression

- Las Máquinas de Vectores Soporte (creadas por **Vladimir Vapnik**) constituyen un método basado en aprendizaje para la resolución de problemas de **clasificación** y **regresión**. Las SVM fueron presentadas en 1992 y adquirieron fama
- cuando dieron resultados muy superiores a las redes neuronales en el reconocimiento de letra manuscrita, usando como entrada pixeles.

El SVM es un algoritmo para encontrar clasificadores lineales

- en **espacios transformados.**

[redacted]

Aplicaciones de las máquinas de vectores de s

- Reconocimiento óptico de caracteres.
- Detección de caras para que las cámaras digitales enfoquen correctamente.
- Filtros de spam para correo electrónico.
- Reconocimiento de imágenes a bordo de satélites (saber qué

partes de una imagen tienen nubes, tierra, agua, hielo, etc.)

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Definición del SVM

Support vector machine (SVM) es un **algoritmo de aprendizaje supervisado** que se utiliza en muchos problemas de **clasificación** y **regresión**, incluidas aplicaciones médicas de procesamiento de señales, procesamiento del lenguaje natural y reconocimiento de imágenes y voz.

Objetivo del algoritmo SVM

El objetivo del algoritmo SVM es encontrar un hiperplano que separe de la mejor forma posible dos clases diferentes de puntos de datos. “De la mejor forma posible” implica el hiperplano con el margen más amplio entre las dos clases.

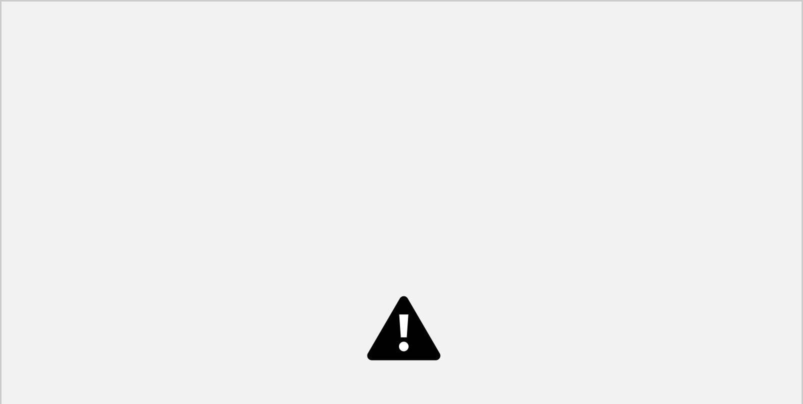


Figura: [redacted]

¿Por qué se llaman Máquinas de Vectores de Soporte?

Se llama «máquina» en español por la parte de «machine». Los vectores de soporte son los puntos que definen el margen máximo de separación del hiperplano que separa las clases. Se llaman vectores, en lugar de puntos, porque estos «puntos» tienen tantos elementos como dimensiones tenga nuestro espacio de entrada.

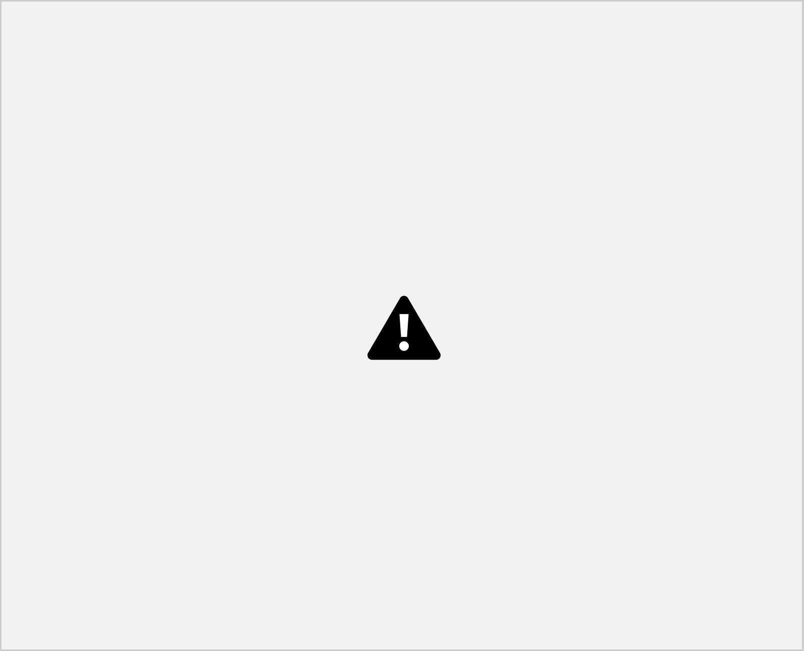
(Data

Figura:

Hiperplano y clasificador de margen máximo

En un espacio p-dimensional, un hiperplano se define como un subespacio plano y afín de dimensiones $p - 1$. El término afín significa que el subespacio no tiene por qué pasar por el origen.

La definición matemática de un hiperplano es bastante simple. Dados los parámetros β_0, $\beta_1, \ldots, \beta_p$ y los pares ($\mathbf{x} = x_1, x_2, \ldots, x_p$) para los que cumple esta igualdad se define la ecuación del hiperplano como sigue:



$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p = 0 \quad (1)$$

Cuando $\mathbf{x}$ no satisface la ecuación se cumple que:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p < 0$$

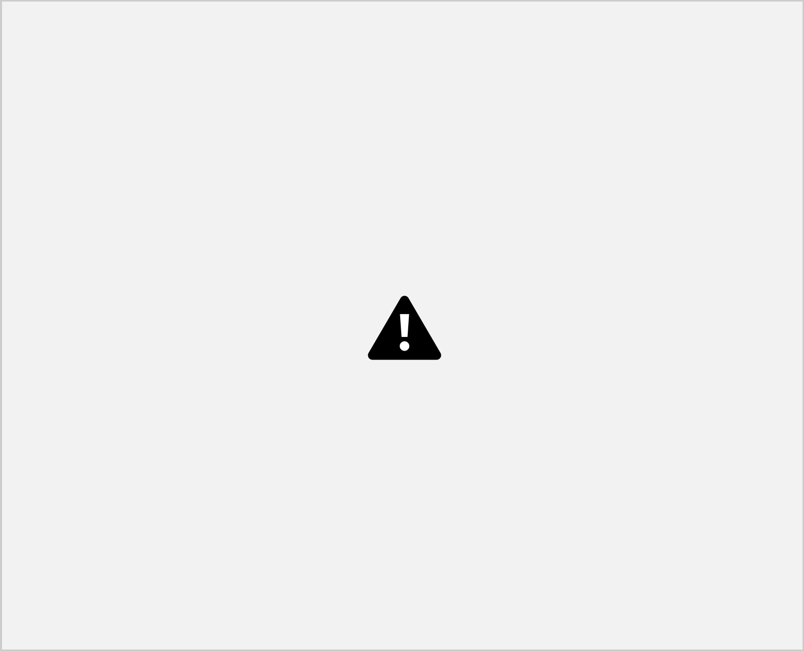
$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p > 0$$

Entonces e puede entender que un hiperplano divide un espacio p-dimensional en dos mitades.

Figura:

Clasificación binaria

Cuando se dispone de *n* observaciones, cada una con *p* predictores y cuya variable respuesta tiene dos categorías, se pueden emplear hiperplanos para construir un clasificador que permita predecir a que grupo pertenece una observación en función de sus predictores.Las siguientes explicaciones se basan en un espacio de dos dimensiones, donde el hiperplano es una recta.



Para casos linealmente separables, la distribución de las observaciones es tal que se pueden separar linealmente de forma perfecta en las dos clases (0 y 1), entonces, un hiperplano de separación cumple que:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p > 0, \; si \; y_i = 1$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p < 0, si\ y_i = 0$$

Clases linealmente separables

La definición de hiperplano para casos perfectamente separables linealmente resulta en un número infinito de posibles hiperplanos, lo que hace necesario un método que permita seleccionar uno de como clasificador óptimo.

Figura:

Selección del clasificador óptimo

La solución a este problema consiste en seleccionar como clasificador óptimo al que se conoce como maximal margin hyperplane. El maximal margin hyperplane se define como el hiperplano que consigue un mayor margen, es decir, que la distancia mínima entre el hiperplano y las observaciones es lo más grande posible.

Figura: [redacted]

¿Como seleccionar el clasificador óptimo?

Para construir el hiperplano de margen máximo (MMH) es como sigue: Dado *n* observaciones de entrenamiento $\vec{x}_1, ..., \vec{x}_n \in R^p$ y asociado a etiquetas de clase, $y_1, ..., y_n \in \{-1, 1\}$, el MMH es la solución al siguiente procedimiento de optimización:

Maximizar $M \in R$, variando $\beta_1, ..., \beta_p$ dado que:

p

[redacted] [redacted]

[redacted]

$$y_i(\vec{\beta} \cdot \vec{x} + \beta_0) \geq M,$$ [redacted]

Clases no separables linealmente

El maximal margin hyperplane descrito en el apartado anterior es una forma muy simple y natural de clasificación siempre y cuando exista un hiperplano de separación. En la gran mayoría de casos reales, los datos no se pueden separar linealmente de forma perfecta, por lo que no existe un hiperplano de separación y no puede obtenerse un maximal margin hyperplane

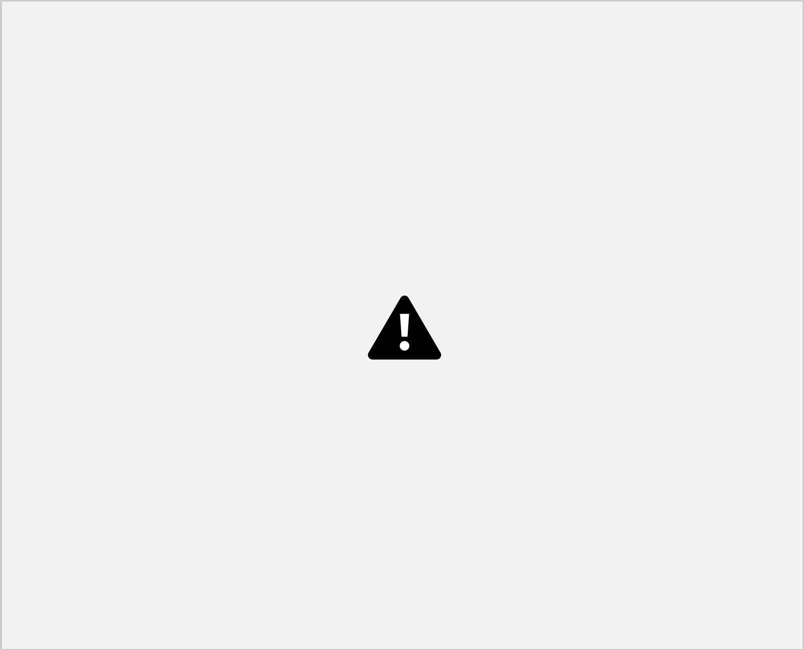


Figura: [redacted]

Clasificadores de vector soporte

El Maximal Margin Classifier descrito en la sección anterior tiene poca aplicación práctica, ya que rara vez se encuentran casos en los que las clases sean perfecta y linealmente separables. Esta aproximación presenta dos inconvenientes:

- Dado que el hiperplano tiene que separar perfectamente las observaciones, es muy sensible a variaciones en los datos. Incluir una nueva observación puede suponer cambios muy grandes en el hiperplano de separación (poca robustez).

- Que el maximal margin hyperplane se ajuste perfectamente a las observaciones de entrenamiento para separarlas todas correctamente suele conllevar problemas de overfitting.

Clasificadores de vector soporte

Por estas razones, es preferible crear un clasificador basado en un hiperplano que, aunque no separe perfectamente las dos clases, sea más robusto y tenga mayor capacidad predictiva al aplicarlo a nuevas observaciones (menos problemas de overfitting). Esto es exactamente lo que consiguen los clasificadores de vector soporte.

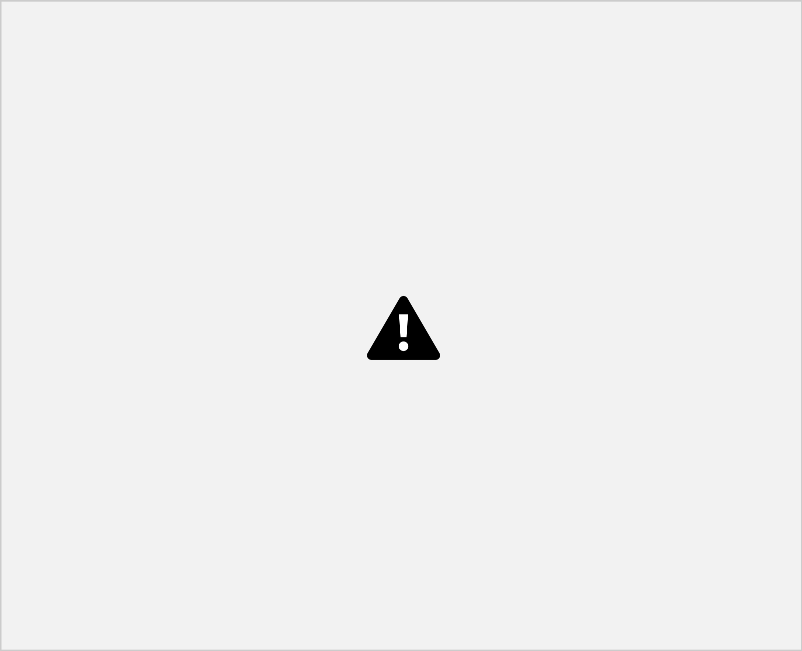


Figura: [redacted]

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Clasificadores de vector soporte

El clasificador de vectores de soporte clasifica una observación de prueba según de qué lado del hiperplano se encuentra. El hiperplano se elige para separar correctamente la mayoría de las observaciones de entrenamiento en las dos clases, pero puede clasificar erróneamente algunas observaciones. El hiperplano elegido es la solución al siguiente problema de optimización: Maximizar M

$\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n$

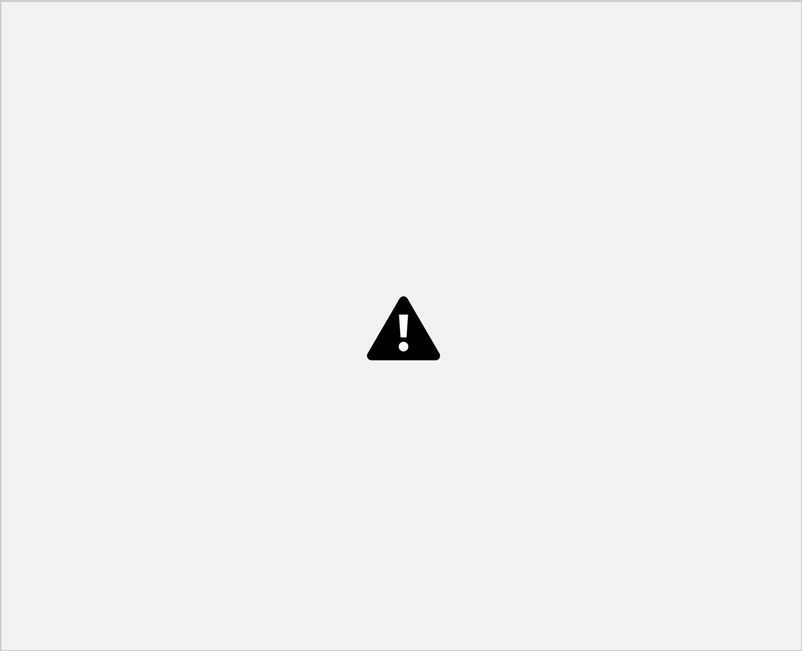
[redacted]

$$y_i\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}\right) \geq M\left(1 - \epsilon_i\right) \quad (8)$$

[redacted]

Hiperparámetro de tuning C

Es importante mencionar que el proceso incluye un hiperparámetro de tuning *C* , el cual controla el número y severidad de las violaciones del margen (y del hiperplano) que se toleran en el proceso de ajuste.



- Cuando *C* = ∞, no se permite ninguna violación del margen y por lo tanto, el resultado es equivalente al Maximal Margin Classifier.
- Cuando *C* se aproxima a cero, menos se penalizan los errores y más observaciones pueden estar en el lado incorrecto del margen o incluso del hiperplano.
- *C* es el hiperparámetro encargado de controlar el balance entre sesgo y varianza del modelo

- En la práctica, su valor óptimo se identifica mediante cross-validation.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Máquinas de Vector Soporte

- El Support Vector Classifier descrito anteriormente consigue buenos resultados cuando el límite de separación entre clases es aproximadamente lineal.
- Si la separación de grupos no es lineal su capacidad ajuste decae drásticamente.
- Una estrategia para enfrentarse a escenarios en los que la separación de los grupos es de tipo no lineal

consiste en expandir las dimensiones del espacio original.

- El método de Máquinas Vector Soporte (SVM) se puede considerar como una extensión del Support Vector Classifier obtenida al aumentar la dimensión de los datos.
- Los límites de separación lineales generados en el espacio aumentado se convierten en límites de separación no lineales al proyectarlos en el espacio original.

Trasformación de espacios

Figura:

Transformación de espacios

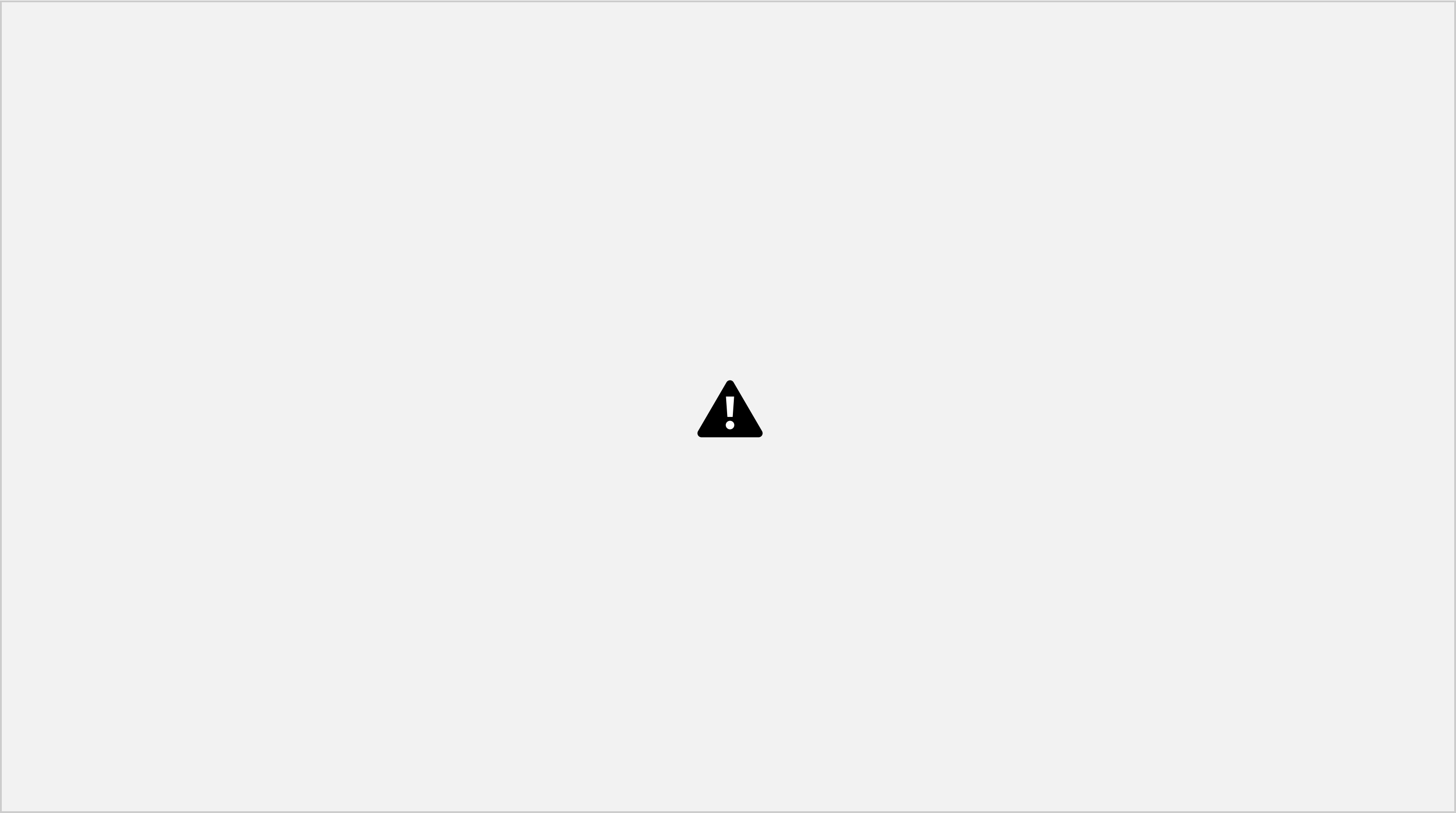


Figura: [redacted]

Transformación de dimensiones - Kernel

Las Máquinas de Vector Soporte siguen la misma estrategía que el Support Vector Classifier, pero aumentando la dimensión de los datos antes de aplicar el algoritmo.

La dimensión de un conjunto de datos puede transformarse combinando o modificando cualquiera de sus dimensiones.

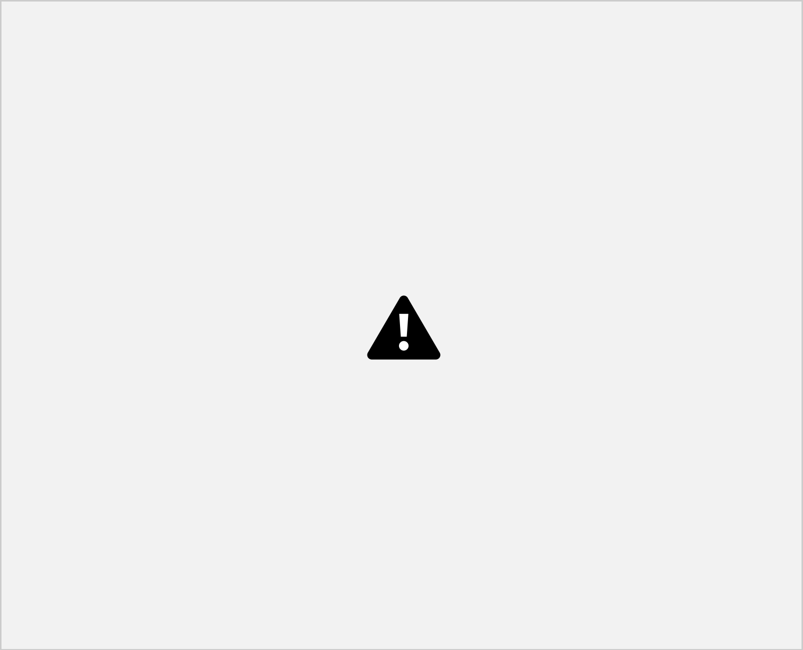
¿Cómo se aumenta la dimensión y qué dimensión es la correcta?

¿Cómo saber cuál es la transformación adecuada?

[redacted]

El truco de Kernel

Existen infinitas transformaciones posibles para la dimensión de un espacio, **¿Cómo saber cuál es la adecuada?** Es aquí donde los kernel entran en juego.



Un *kernel* (K) es una función que devuelve el resultado del producto interno entre dos vectores en un nuevo espacio dimensional distinto al espacio original en el que se encuentran los vectores.

- Si se sustituye este *producto interno* por un kernel, se obtienen directamente los vectores soporte (y el hiperplano) en la dimensión correspondiente al kernel.
- La utilidad de los kernels radica en que se puede obtener el resultado para cualquier dimensión.

Tipos de Kernel

1. Kernel Lineal

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}' \quad (10)$$

Figura: [redacted]

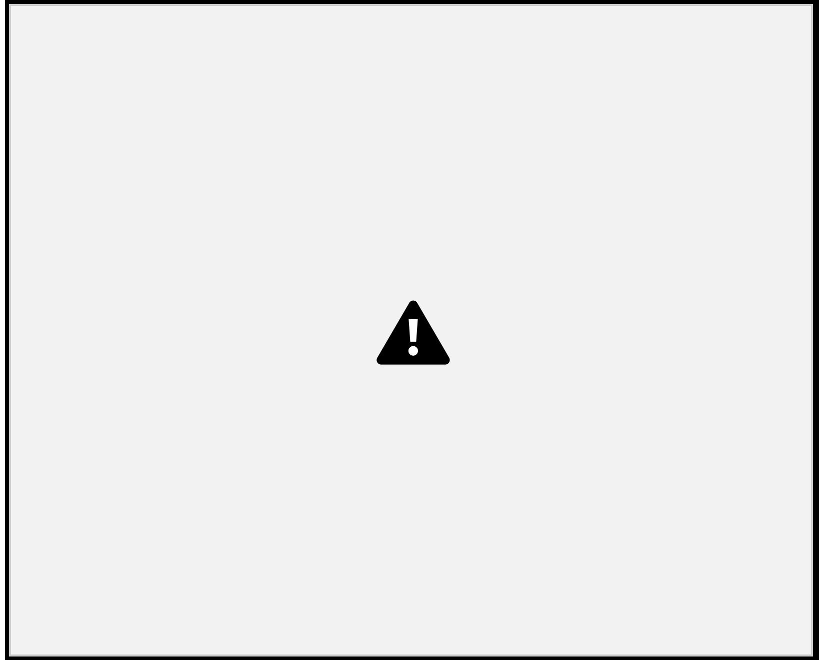
Tipos de Kernel

[redacted]

d

$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + c)$ [redacted]

[redacted]



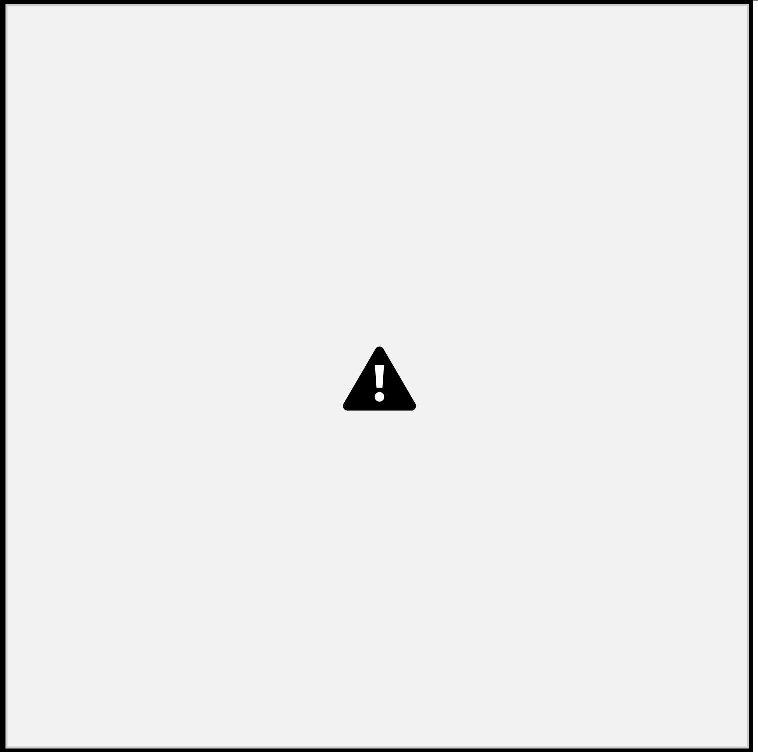
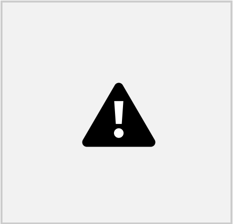


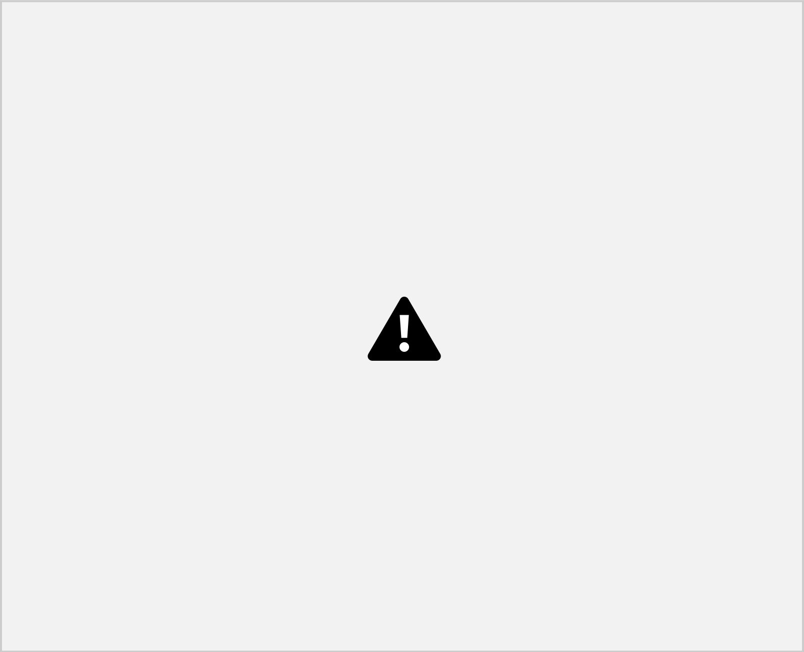
Figura: [redacted]

Tipos de Kernel



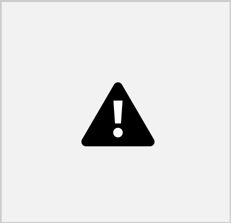
3. Kernel Gaussiano (RBF)

El valor de γ controla el comportamiento del kernel, cuando es muy pequeño, el modelo final es equivalente al obtenido con un kernel lineal, a medida que aumenta su valor, también lo hace la flexibilidad del modelo.



$$K(\mathbf{x}, \mathbf{x}') = exp(-\gamma||\mathbf{x} - \mathbf{x}'||^2) \quad (12)$$

Recomen[illegible]Kernel



- Los kernels tienen una serie de hiperparámetros cuyo valor óptimo puede encontrarse mediante validación cruzada.
- No existe un kernel que supere al resto, depende en gran medida de la naturaleza del problema que se esté estudiando.
- Muchos autores recomiendan probar el kernel RBF. Este kernel tiene dos ventajas: que solo tiene dos hiperparámetros que optimizar (γ y la penalización C común a todos los SVM) y que su flexibilidad puede ir desde un clasificador lineal a uno muy complejo.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com Machine Learning

Referencias

[redacted] [redacted]

[redacted] [redacted]