

Máquina de Soporte Vectorial (SVM)

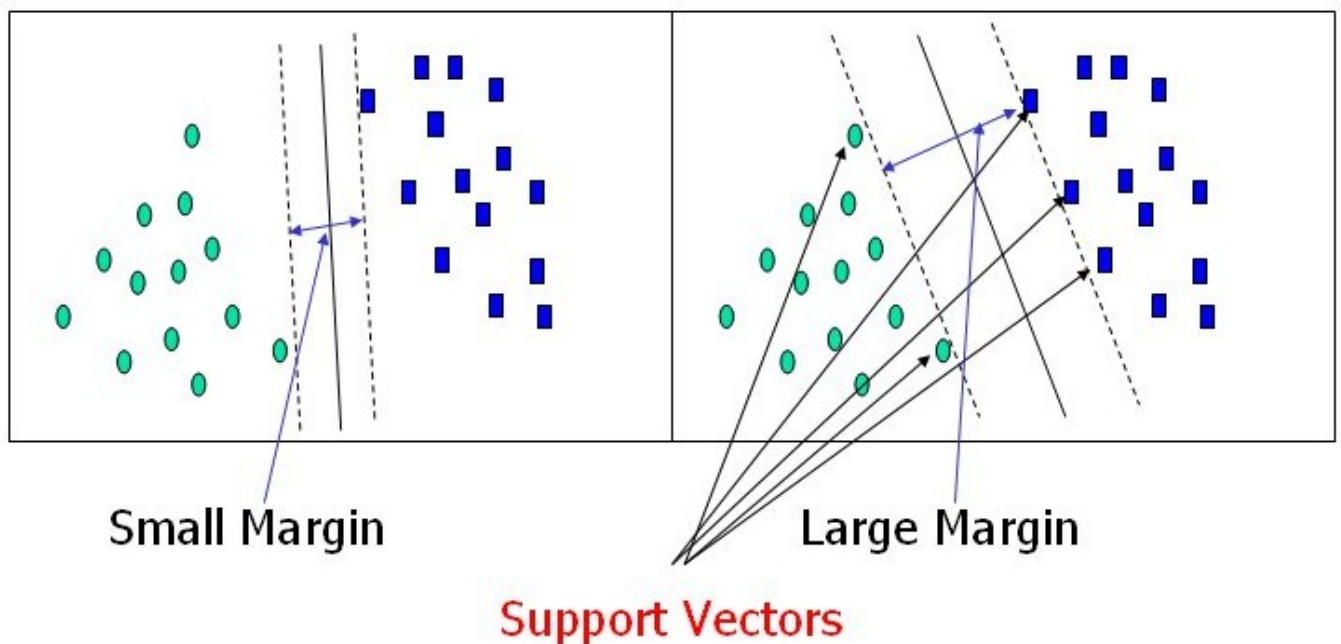


Dado un conjunto de ejemplos de entrenamiento, cada uno marcado como perteneciente a una u otra de dos categorías, un algoritmo de entrenamiento de SVM construye un modelo que asigna nuevos ejemplos a una categoría u otra, convirtiéndolo en un clasificador lineal binario no probabilístico. ([Wikipedia](#))

Entonces, ¿Qué es una máquina de vectores de soporte?

Es un algoritmo de aprendizaje automático supervisado que se puede utilizar para problemas de clasificación o regresión. Pero generalmente se usa para clasificar. Dadas 2 o más clases de datos etiquetadas, actúa como un clasificador discriminativo, definido formalmente por un hiperplano óptimo que separa todas las clases. Los nuevos ejemplos que luego se mapean en ese mismo espacio se pueden clasificar según el lado de la brecha en que se encuentran.

Como regla general, las SVM son excelentes para conjuntos de datos relativamente pequeños con menos valores atípicos.



Máquina de Soporte Vectorial (SVM)



Representación de los vectores de soporte.

Los vectores de soporte son los puntos de datos más cercanos al hiperplano, los puntos de un conjunto de datos que, si se eliminan, alterarían la posición del hiperplano en división. Debido a esto, pueden considerarse los elementos críticos de un conjunto de datos, son los que nos ayudan a construir nuestra SVM.

Ahora, ¿Qué es un hiperplano?

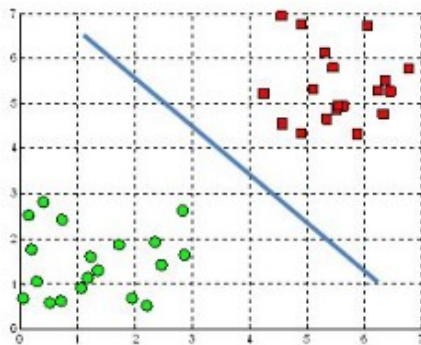
La geometría nos dice que un hiperplano es un subespacio de una dimensión menos que su espacio ambiental. Por ejemplo, un hiperplano de un espacio n -dimensional es un subconjunto plano con dimensión $n - 1$. Por su naturaleza, separa el espacio en dos medios espacios.

Máquina de Soporte Vectorial (SVM)

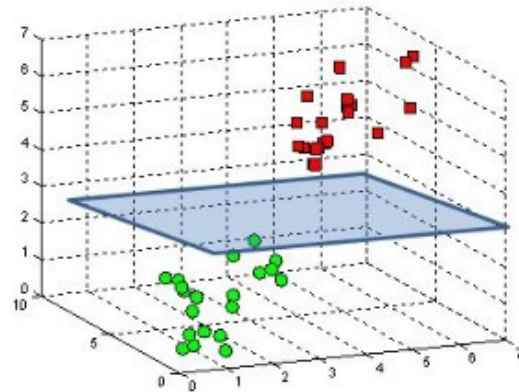
Hyperplanes as decision surfaces

- A hyperplane is a linear decision surface that splits the space into two parts;
- It is obvious that a hyperplane is a binary classifier.

A hyperplane in \mathbb{R}^2 is a line



A hyperplane in \mathbb{R}^3 is a plane



A hyperplane in \mathbb{R}^n is an $n-1$ dimensional subspace

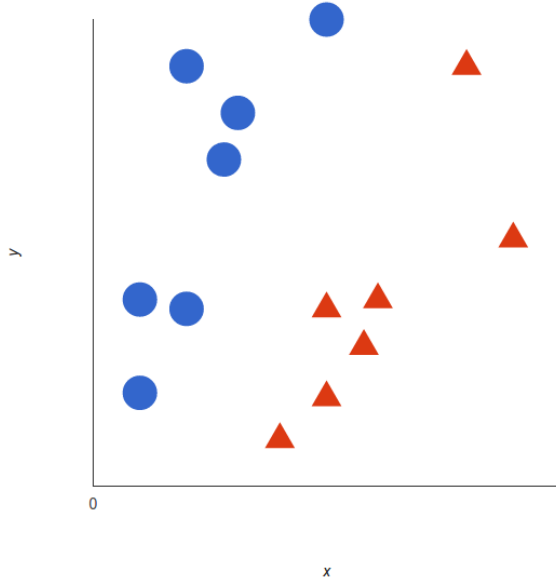
32

Hiperplanos como desicion de superficies.

¿Cómo funciona SMV?

Los conceptos básicos de Support Vector Machines y su funcionamiento se comprenden mejor con un ejemplo sencillo. Imaginemos que tenemos dos etiquetas: rojos y azules , y nuestros datos tiene dos características : x e y . Queremos un clasificador que, dado un par de coordenadas (x, y) , dé como resultado si es rojo o azul . Trazamos nuestros datos de entrenamiento ya etiquetados en un avión:

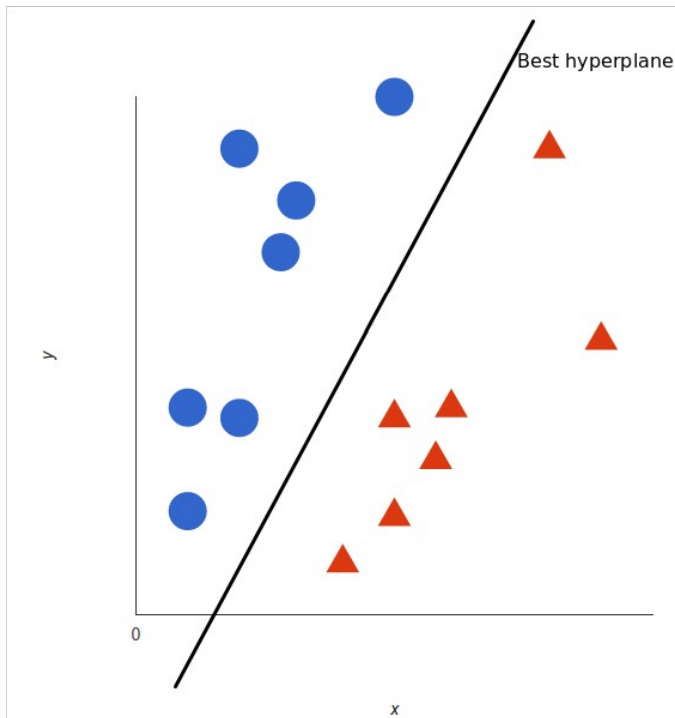
Máquina de Soporte Vectorial (SVM)



Datos etiquetados

Una máquina de vectores de soporte toma estos puntos de datos y genera el hiperplano (que en dos dimensiones es simplemente una línea) que separa mejor las etiquetas. Esta línea es el límite de decisión : todo lo que caiga a un lado lo clasificaremos como azul y todo lo que caiga al otro lado como rojo .

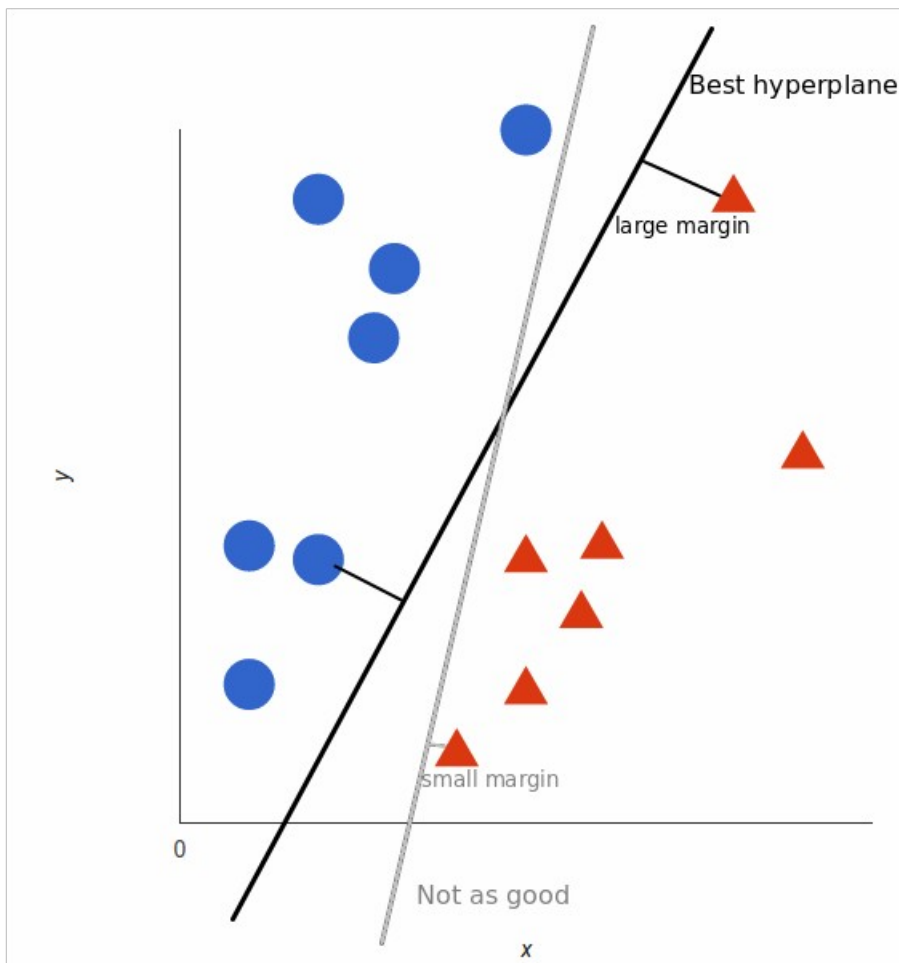
Máquina de Soporte Vectorial (SVM)



Como vimos, en 2D el hiperplano es simplemente una línea.

Pero, ¿cuál es exactamente el mejor hiperplano? Para SVM, es el que maximiza los márgenes de ambas etiquetas. En otras palabras: el hiperplano (recuerde que es una línea en este caso) cuya distancia al elemento más cercano de cada etiqueta es la más grande.

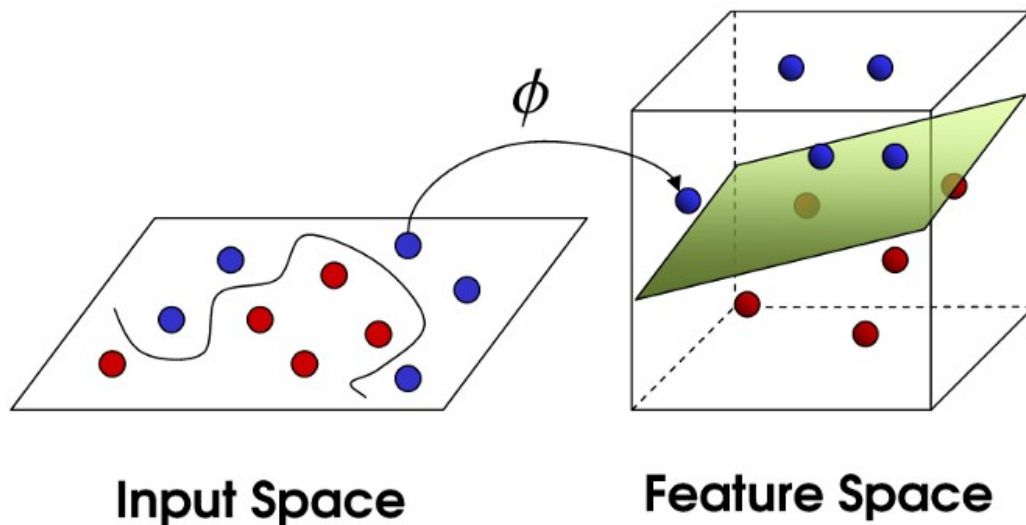
Máquina de Soporte Vectorial (SVM)



No todos los hiperplanos son iguales.

Truco del Kernel

Máquina de Soporte Vectorial (SVM)



Representación del truco del kernel.

El truco del kernel que le permite ampliar los datos en infinitas dimensiones.

Ciertamente no podré explicar completamente el truco del kernel en esta publicación. Creo que una comprensión verdaderamente profunda del truco del kernel requiere un tratamiento matemático riguroso, y esto no se puede hacer en lo que se etiqueta como una lectura de siete minutos, y mucho menos por alguien que de manera realista solo ha estado aprendiendo sobre el truco del kernel y las SVM para un corto período de tiempo.

Definamos nuestra función de pérdida (qué minimizar)

Usaremos Hinge Loss. Esta es una función de pérdida utilizada para entrenar clasificadores. La Hinge Loss se utiliza para la clasificación de “margen máximo”.

$$c(x, y, f(x)) = (1 - y * f(x))_+$$

c es la función de pérdida, x la muestra, y es la etiqueta verdadera, $f(x)$ la etiqueta predicha.

Máquina de Soporte Vectorial (SVM)



Definamos nuestra función objetivo (qué optimizar)

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)_+$$

La función objetivo de una SVM consta de dos términos. El primer término es un regularizador, el segundo término es la pérdida.

1. El regularizador equilibra la maximización del margen y la pérdida.
2. Queremos encontrar la superficie de decisión que esté lo más lejos posible de cualquier punto de datos.

Entonces, ¿Cómo podemos minimizar nuestra pérdida y optimizar para nuestro objetivo?

Tenemos que derivar nuestra función objetivo para obtener los gradientes.

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} (1 - y_i \langle x_i, w \rangle)_+ = \begin{cases} 0, & \text{if } y_i \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases}$$

Esto significa que, si tenemos una muestra mal clasificada, actualizamos el vector de peso w usando los gradientes de ambos términos; de lo contrario, si se clasifica correctamente, simplemente actualizamos w por el gradiente del regularizador.

Actualizar la regla para w (cuando esta mal clasificada)

$$w = w + \eta(y_i x_i - 2\lambda w)$$

La tasa de aprendizaje es la longitud de los pasos que el algoritmo realiza por el gradiente en la curva de error.

El regularizador controla la compensación entre lograr un error de entrenamiento bajo y un error de prueba bajo que es la capacidad de generalizar su clasificador a datos invisibles.

Máquina de Soporte Vectorial (SVM)



Actualizar la regla para w (cuando esta bien clasificada)

$$w = w + \eta(-2\lambda w)$$

Actualizar el valor de w .

Conclusiones

Lo que hace una SVM es averiguar la “mejor” línea para dividir los puntos, que es la línea que deja la mayor brecha entre ella y los puntos. Esa línea se llama hiperplano de margen máximo, porque es un hiperplano que separa los lados dejando el margen máximo.

también resulta que la mayoría de los puntos realmente no importan cuando averiguas dónde colocar la línea, porque están muy lejos de ella. Solo los puntos cercanos importan, y debido a que “apoyan” la ubicación de la línea, los llamamos vectores de soporte. Gracias a la matemática, ni siquiera necesita calcular dónde está esta línea / plano para averiguar de qué lado están las cosas; todo lo que necesita son estos vectores de soporte.

Hay incluso más beneficios al usar solo los vectores de soporte para construir una SVM; por ejemplo, ¿qué sucede si sus puntos no se pueden dividir claramente con una línea? Hay algo que se llama el truco del kernel que le permite ampliar los datos en infinitas dimensiones, donde si hay alguna curva de separarlos, que se encuentran sus vectores de soporte el cual es literalmente imposible si desea construir un SVM solo calculando el hiperplano de separación directamente.