

# Algoritmos esenciales de ML

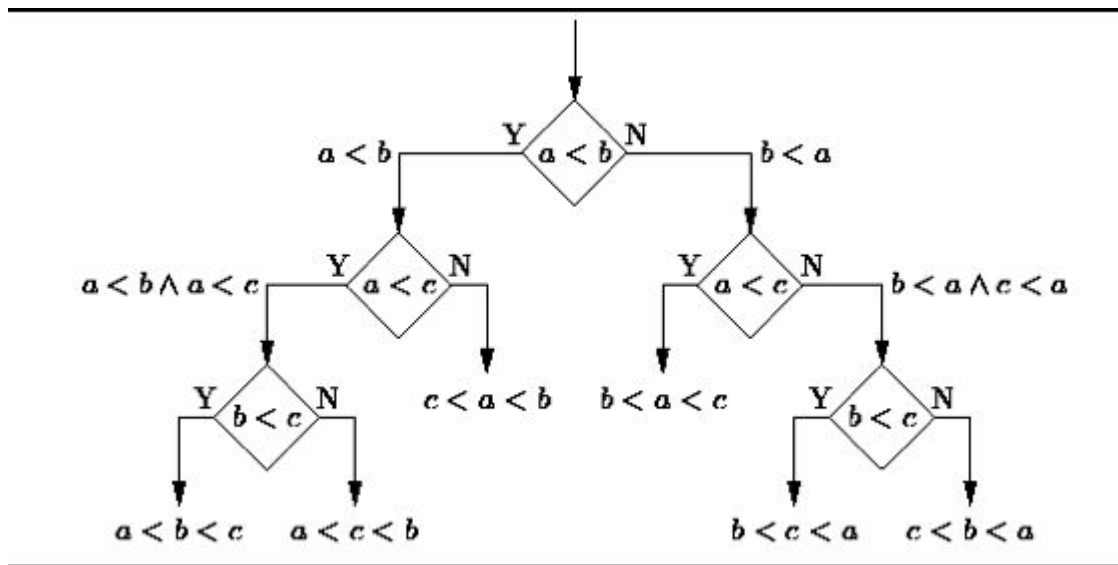
# Árboles de decisión

Un árbol de decisiones es una herramienta de apoyo a la decisión que utiliza un gráfico que considera las posibles decisiones y sus posibles consecuencias, incluidos los resultados de eventos fortuitos, los costos de recursos y la utilidad.

Desde el punto de vista de la toma de decisiones empresariales, un árbol de decisiones es el número mínimo de preguntas si/no que uno tiene que hacer, para evaluar la probabilidad de tomar una decisión correcta, la mayoría del tiempo.

Este método permite abordar el problema de una manera estructurada y sistemática para llegar a una conclusión lógica.

# Árboles de decisión

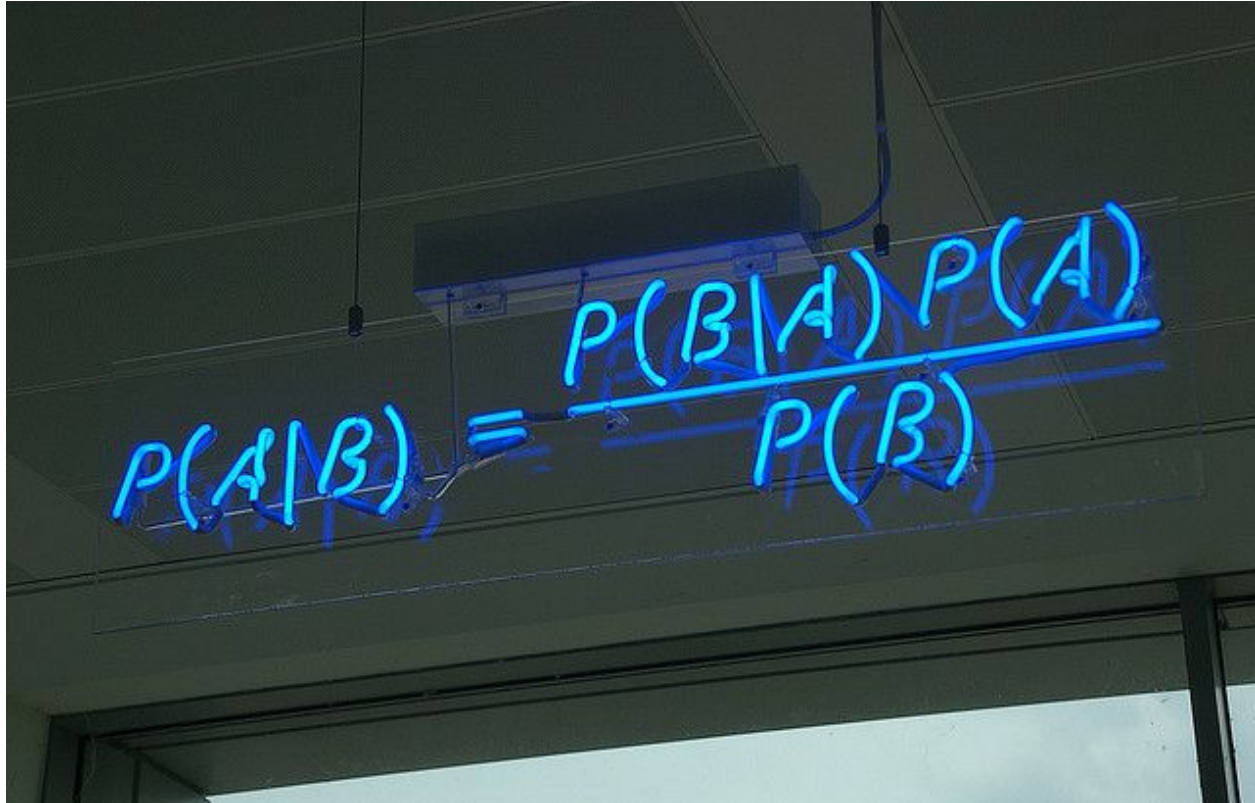


# Naïve Bayes Clasification

Los clasificadores Naïve Bayes son una familia de simples clasificadores probabilísticos basado en la aplicación de Bayes 'teorema con fuertes (Naïve) supuestos de independencia entre las características'.

La imagen destacada es la ecuación – con  $P(A | B)$  es probabilidad posterior,  $P(B | A)$  es probabilidad,  $P(A)$  es probabilidad previa de clase, y  $P(B)$  predictor probabilidad previa.

# Naïve Bayes Clasification



A photograph of a blue neon sign mounted on a dark ceiling. The sign displays the Naïve Bayes classification formula in a stylized, handwritten font. The formula is  $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ . The sign is illuminated with a bright blue light, and the background is dark, making the neon text stand out. The sign is slightly tilted and has some visible wiring and mounting hardware.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

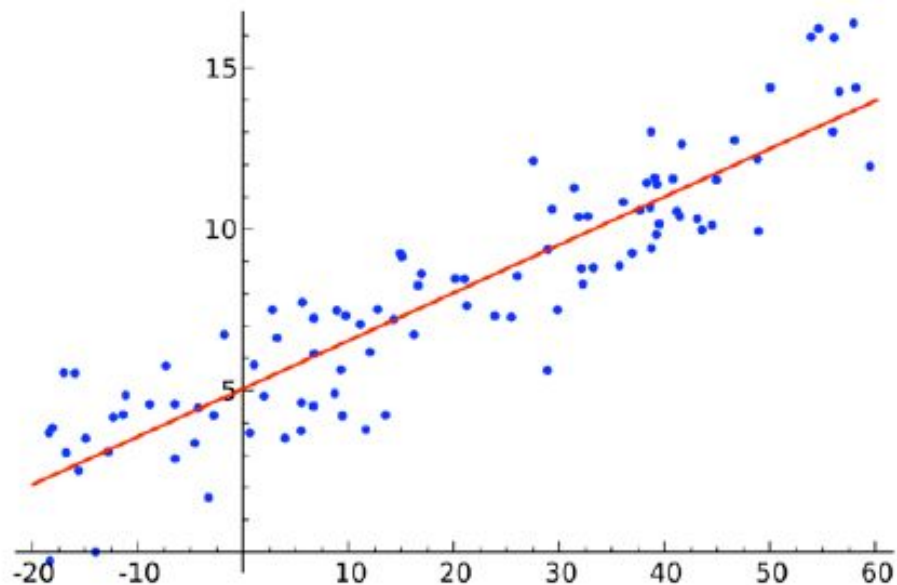
# Regresión Lineal

Se puede pensar en la regresión lineal como la tarea de ajustar una línea recta a través de un conjunto de puntos.

Hay varias estrategias posibles para hacer esto, y la estrategia de “mínimos cuadrados ordinarios” va así: puede dibujar una línea y luego, para cada uno de los puntos de datos, medir la distancia vertical entre el punto y la línea y sumarlos; La línea ajustada sería aquella en la que esta suma de distancias sea lo más pequeña posible.

Linear se refiere al tipo de modelo que está utilizando para ajustar los datos, mientras que los mínimos cuadrados se refieren al tipo de métrica de error que está minimizando.

# Regresión Lineal



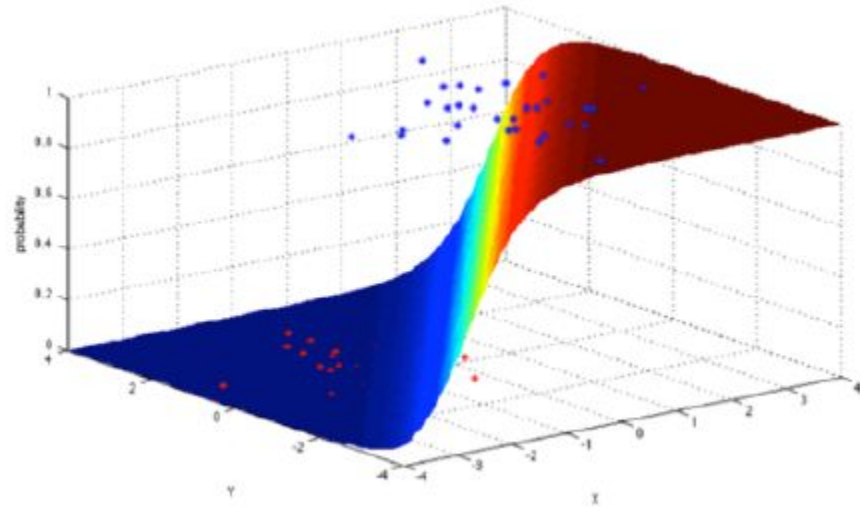
# Regresión Logística

La regresión logística es una poderosa manera estadística de modelar un resultado binomial con una o más variables explicativas.

Mide la relación entre la variable dependiente categórica y una o más variables independientes estimando las probabilidades utilizando una función logística, que es la distribución logística acumulativa.



# Regresión Logística

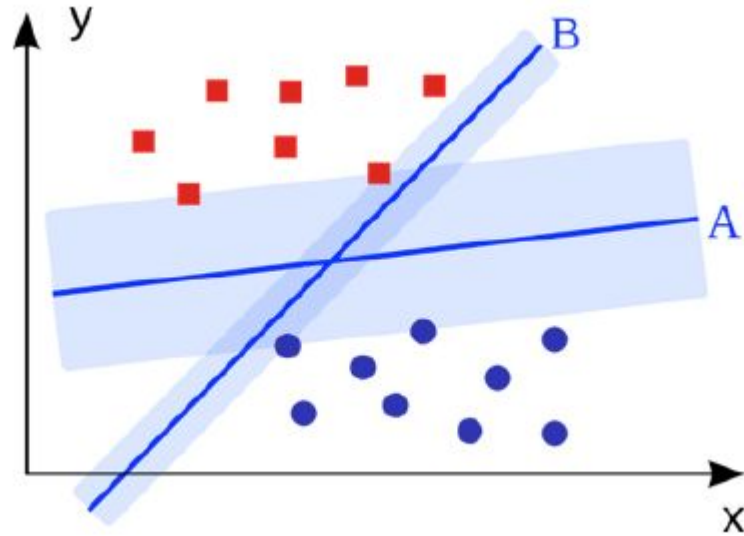


# Support Vector Machines

SVM es un algoritmo de clasificación binario. Dado un conjunto de puntos de 2 tipos en el lugar  $N$  dimensional, SVM genera un hiperplano  $(N-1)$  dimensional para separar esos puntos en 2 grupos. Digamos que se tiene algunos puntos de 2 tipos en un papel que son linealmente separables. SVM encontrará una línea recta que separa esos puntos en 2 tipos y situados lo más lejos posible de todos esos puntos.

En términos de escala, algunos de los mayores problemas que se han resuelto utilizando SVMs (con implementaciones adecuadamente modificadas) son publicidad en pantalla, reconocimiento de sitios de empalme humanos, detección de género basada en imágenes, clasificación de imágenes a gran escala.

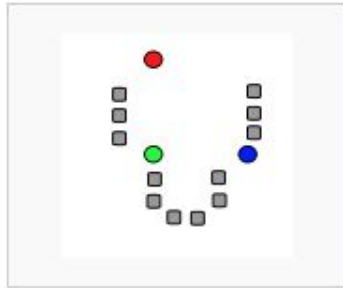
# Support Vector Machines



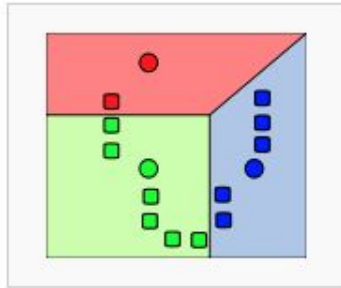
# Algoritmos Clustering

Clustering es la tarea de agrupar un conjunto de objetos tales que los objetos en el mismo grupo (cluster) son más similares entre sí que a los de otros grupos.

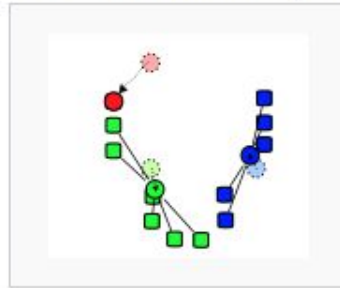
## Demostración del algoritmos estándar



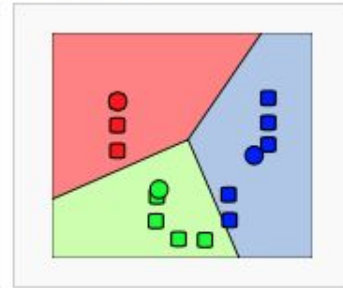
1)  $k$  centroides iniciales (en este caso  $k=3$ ) son generados aleatoriamente dentro de un conjunto de datos (mostrados en color).



2)  $k$  grupos son generados asociándole el punto con la media más cercana. La partición aquí representa el [diagrama de Voronoi](#) generado por los centroides.

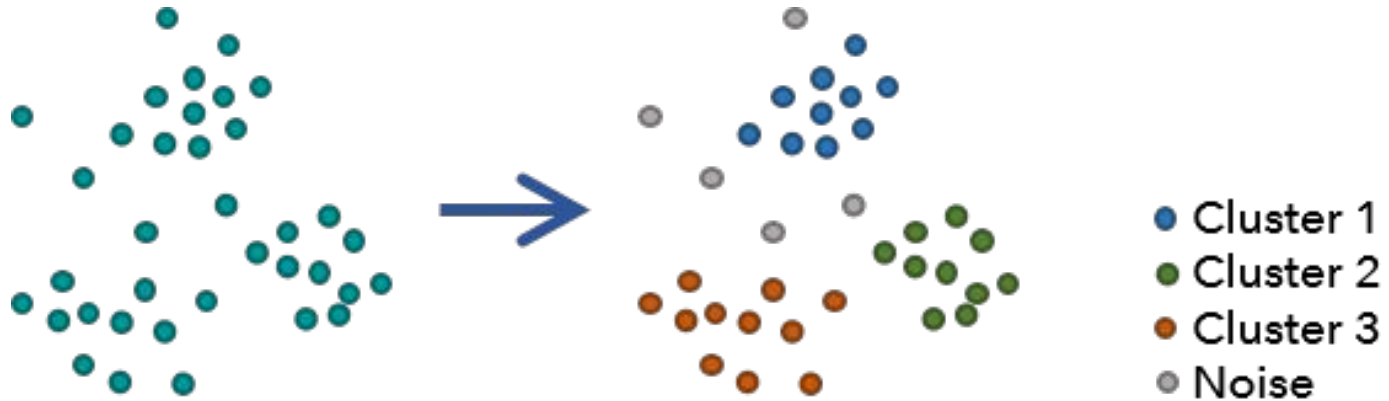


3) EL [centroide](#) de cada uno de los  $k$  grupos se recalcula.



4) Pasos 2 y 3 se repiten hasta que se logre la convergencia.

# Algoritmos Clustering



# Análisis de componentes Principales PCA

PCA es un procedimiento estadístico que usa una transformación ortogonal para convertir un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables linealmente no correlacionadas llamadas componentes principales.

Algunas de las aplicaciones de PCA incluyen compresión, simplificación de datos para un aprendizaje más fácil, visualización. No es adecuado en los casos en que los datos son ruidosos.

# Análisis de componentes Principales PCA

