

Modulo 5: Machine Learning en Python

En este módulo vamos a hablar de algunas técnicas de machine learning en Python, como la regresión lineal y las máquinas de soporte vectorial.

Técnicas de machine learning:

Las técnicas más conocidas del machine learning, son la regresión, la clasificación y el clustering.

Regresión lineal simple:

La regresión lineal busca relacionar dos variables por medio de una línea recta y permite predecir el valor de una variable si conocemos la otra. Por ejemplo, si se tienen datos de los beneficios monetarios generados por varios valores de inversión, se puede representar una gráfica similar a la mostrada en la ilustración 1, en donde se puede relacionar el beneficio que genera una inversión de cierto valor, la línea naranja representa la relación existente y conociendo la ecuación de esta recta se puede predecir el valor del beneficio de una inversión aún sin conocer exactamente ese dato.

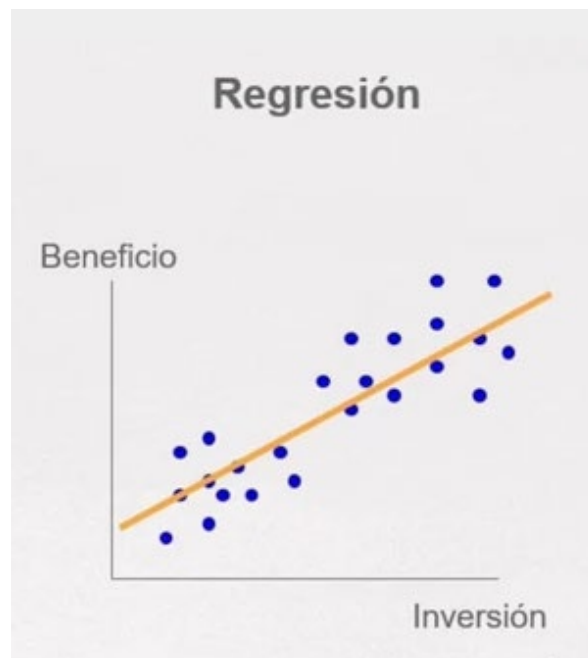


Ilustración 1. Regresión lineal. Fuente: <https://openwebinars.net/blog/modelos-de-machine-learning/>

Una recta tiene esta ecuación característica:

$$y = mx + b$$

En dónde y representa en este ejemplo el valor del beneficio, m que tan pendiente está la recta, x representa el valor de la inversión, y b representa el intercepto de la recta con el eje y .

Los algoritmos de regresión lineal consisten en analizar los datos y sacar un modelo predictivo que permita obtener el valor de una variable si conocemos la otra, como por ejemplo el modelo de regresión lineal permite asociarlos mediante la ecuación de una línea recta donde el modelo entrega los coeficientes para poder calcular la variable de interés.

Técnicas de clasificación

Consisten en encontrar un límite que separe, lo mejor posible, los valores de dos o más variables (este último conocido como clasificación multivariada), para clasificar o decir si un nuevo dato pertenece a una categoría u otra. Por ejemplo, en la ilustración 2 vemos dos variables como la estatura y el peso, y aparte tenemos unas categorías si es niño o adulto (donde las x son niños y los círculos son adultos). El modelo de clasificación se debe entrenar con unos datos ya conocidos, para eso lo que se hace es separar los datos en dos subconjuntos, como se observa en la ilustración 3, un conjunto de entrenamiento, con los cuales el modelo va a definir los parámetros de clasificación y un conjunto de prueba, en donde se va a verificar la precisión del modelo.

Tanto la de regresión como la clasificación son técnicas de aprendizaje supervisado, es decir, que necesitan ser entrenados para poder establecer modelos válidos.

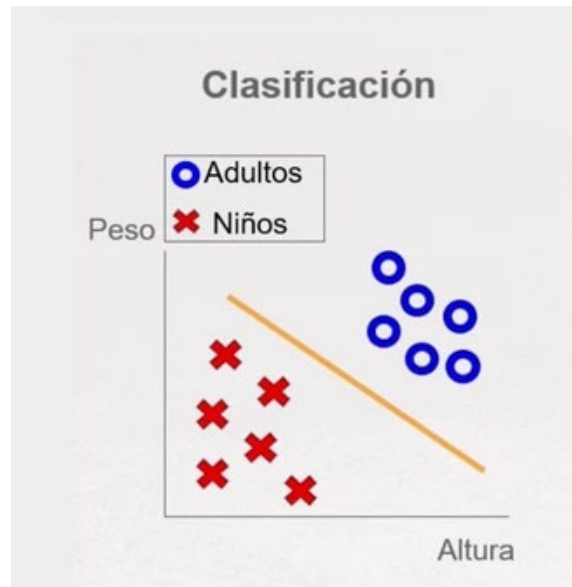


Ilustración 2. Clasificadores. Fuente: <https://openwebinars.net/blog/modelos-de-machine-learning/>

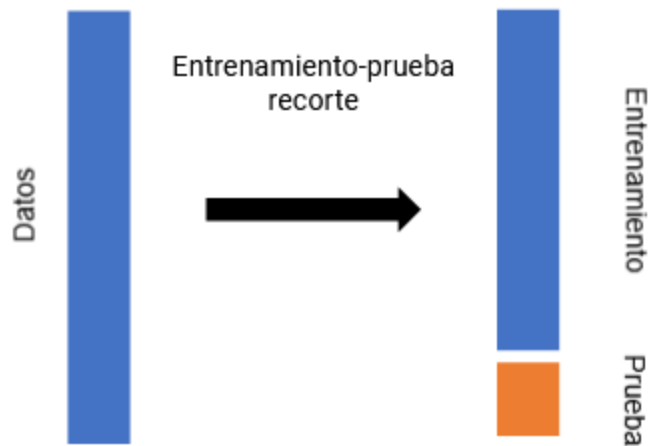


Ilustración 3. Conjunto de entrenamiento y prueba.

Máquinas de soporte vectorial:

Las máquinas de soporte vectorial son algoritmos de clasificación y tiene variedad de aplicaciones tales como detección de rostros, detección de intrusos, clasificación de correos electrónicos, artículos de noticias y páginas web, entre otros. El SVM, por sus siglas en inglés, lo que busca es construir un plano máximo para separar las diferentes clases, por ejemplo, como vemos en la ilustración 4 tenemos la clase A y la clase B, en donde las máquinas de soporte vectorial buscan encontrar un plano que los separe pero que esté a máxima distancia de los datos que están en el borde, que se conocen como soportes vectoriales. En el ejemplo vemos que la separación es fácil ya que es una línea recta, pero con la mayoría de los datos no es tan ideal. una de las mayores ventajas de los SVM, son que se pueden usar con diferentes kernels, ellos le permiten a las SVM buscar planos de separación que no sean necesariamente lineales, en la imagen vemos algunos ejemplos. Lineales, RBF, polinómicos, como se observan en la ilustración 5 y permiten mejorar el plano de separación y la clasificación del modelo.

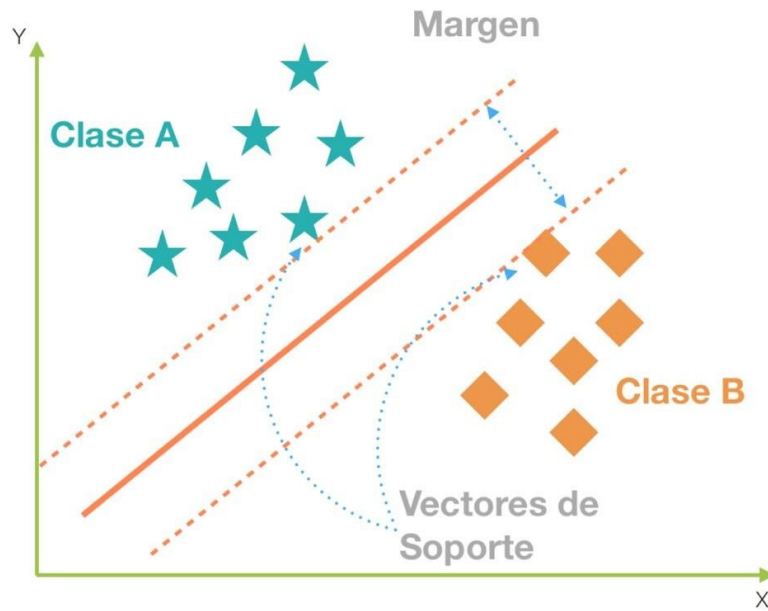


Ilustración 4. Máquina de soporte vectorial. Fuente: <https://aprendeia.com/maquinas-vectores-de-soporte-clasificacion-teoria/>

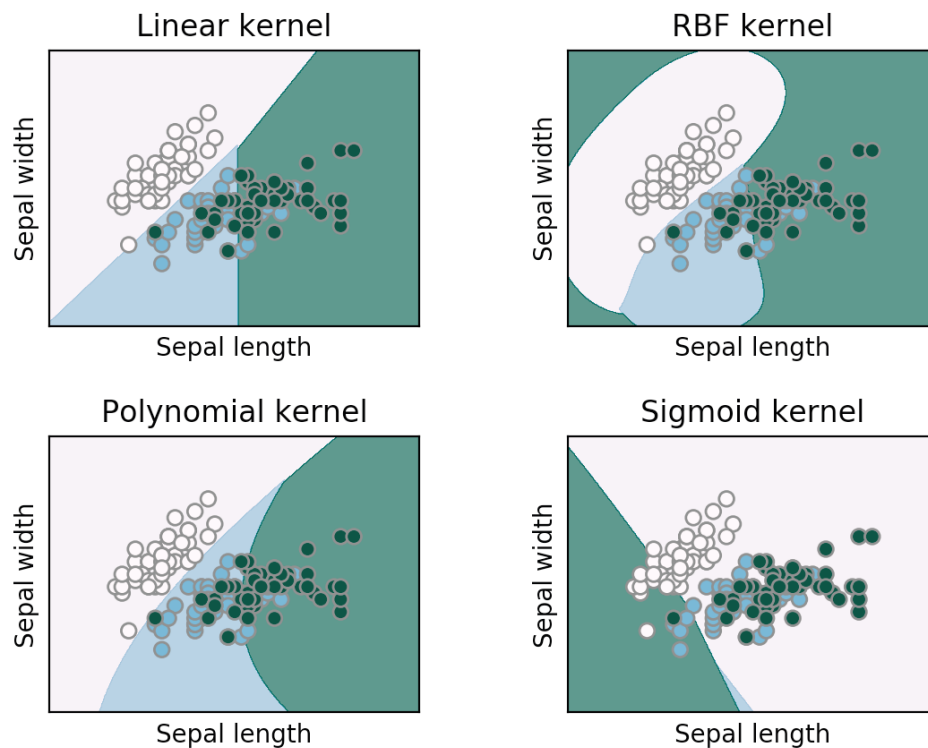


Ilustración 5. Tipos de kernel. Fuente: <https://towardsdatascience.com/multiclass-classification-with-support-vector-machines-svm-kernel-trick-kernel-functions-f9d5377d6f02>