¿Cómo funcionan los árboles de decisiones?

El árbol comienza con un nodo raiz que representa el conjunto completo de datos el cual es porcionado en nodos hijos dada una caracteristica en particular del conjunto de datos.

¿Cuál es la diferencia entre un árbol de decisión y un bosque aleatorio?

Un árbol de decision es un solo árbol que se construye a partir de un conjunto de datos, mientras que el bosque aleatorio es una coleccion de arboles de decision construidos a partir de diferentes subconjuntos de datos y caracteristicas.

¿Qué es bootstrapping?

Es una tecnica que se utiliza para estimar la distribucion de una estadistica de interes a partir de un conjunto de datos. Esta se basa en crear multiples muestras de datos a partir de la muestra original mediante el muestreo aleatorio con reemplazo.

¿Qué es bagging?

Es la combinacion de bootstrap y aggregating.

Se basa en la creacion de multiples conjuntos de datos de entrenamiento a partir de la muestra original mediante el muestreo aleatorio con reemplazo, bootstrapping, y luego entrenar un modelo separado en cada conjunto de datos.

¿Qué son los datos de entrenamiento y para qué se usan?

Son un conjunto de datos utilizados para entrenar un modelo de machine learning.

¿Qué es un conjunto de prueba y por qué usarlo?

Son un conjunto de datos utilizados para aplicar el proceso de entrenamiento a un conjunto “ nuevo” simulando una prediccion en data nueva.

¿Cuáles son los hiperparámetros principales que pueden ajustar para los árboles de decisiones?

Profundidad maxima del árbol, max\_depth.

¿Cuáles son algunas maneras de reducir el sobreajuste con los árboles de decisiones?

El sobreajuste es cuando los modelos se ajustan demasiado bien a los datos de entrenamiento y tienen un rendimiento deficiente en datos nuevos. Para esto se puede:

Limitar la profundidad del árbol.

Aumentar las caracteristicas del conjunto de datos.