Instituto Tecnológico de Tijuana

Nombre de Facultad:

Ingeniería Informática

Proyecto / Tarea / Práctica:

Decision Tree Classifier

Materia:

Datos Masivos

Facilitador:

Jose Christian Romero Hernandez

Alumnos:

Erik Saul Rivera Reyes

Brayan Baltazar Moreno

Alonso Villegas Luis Antonio

Rafael Sanchez Baez

Fecha:

Tijuana Baja California a 07 de 04 2022

Árboles de decisión

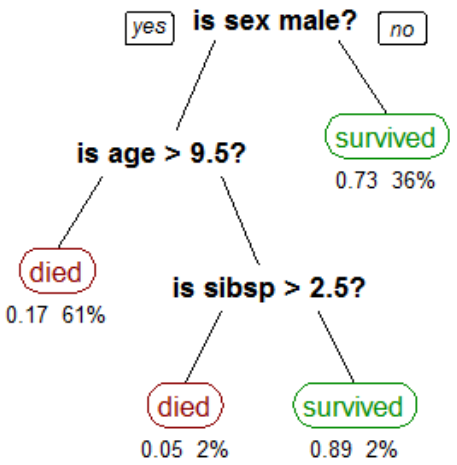
Los árboles de decisión y sus conjuntos son métodos populares para las tareas de clasificación y regresión del aprendizaje automático. Los árboles de decisión se usan ampliamente porque son fáciles de interpretar, manejan características categóricas, se extienden a la configuración de clasificación multiclase, no requieren escalado de características y pueden capturar no linealidades e interacciones de características. Los algoritmos de conjuntos de árboles, como los bosques aleatorios y el impulso, se encuentran entre los mejores para las tareas de clasificación y regresión.

Un árbol de decisión tiene una estructura similar a un diagrama de flujo donde un nodo interno representa una característica o atributo, la rama representa una regla de decisión y cada nodo u hoja representa el resultado. El nodo superior de un árbol de decisión se conoce como nodo raíz.

La idea básica detrás de cualquier problema de árbol de decisión es la siguiente:

- Selecciona el mejor atributo utilizando una medida de selección de atributos o características.
- Haz de ese atributo un nodo de decisión y divide el conjunto de datos en subconjuntos más pequeños.
- Comienza la construcción del árbol repitiendo este proceso recursivamente para cada atributo hasta que una de las siguientes condiciones coincida:
    - Todas las variables pertenecen al mismo valor de atributo.
    - Ya no quedan más atributos.
    - No hay más casos.

Aprendizaje basado en árboles de decisión es un método comúnmente utilizado en la minería de datos.El objetivo es crear un modelo que predice el valor de una variable de destino en función de diversas variables de entrada.



Un árbol de decisión es una representación simple para clasificar ejemplos. El aprendizaje basado en árboles de decisión es una de las técnicas más eficaces para la clasificación supervisada. Para esta sección, se supone que todas las funciones tienen dominios discretos finitos, y existe una sola característica de destino llamado la clasificación. Cada elemento del dominio de la clasificación se llama clase. Un árbol de decisión o un árbol de clasificación es un árbol en el que cada nodo interno (no hoja) está etiquetado con una función de entrada. Los arcos procedentes de un nodo etiquetado con una característica están etiquetados con cada uno de los posibles valores de la característica. Cada hoja del árbol se marca con una clase o una distribución de probabilidad sobre las clases.

Un árbol puede ser "aprendido" mediante el fraccionamiento del conjunto inicial en subconjuntos basados en una prueba de valor de atributo. Este proceso se repite en cada subconjunto derivado de una manera recursiva llamada particionamiento recursivo. La recursividad termina cuando el subconjunto en un nodo tiene todo el mismo valor de la variable objetivo, o cuando la partición ya no agrega valor a las predicciones. Este proceso de inducción top-down de los árboles de decisión es un ejemplo de un algoritmo voraz, y es, con mucho, la estrategia más común para aprender árboles de decisión a partir de datos.

En minería de datos, los árboles de decisión se pueden describir también como la combinación de técnicas matemáticas y computacionales para ayudar a la descripción, la categorización y la generalización de un conjunto dado de datos.

Los datos provienen en registros de la forma:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \ldots, x_k, Y)$$

La variable dependiente, Y, es la variable objetivo que estamos tratando de entender, clasificar o generalizar. El vector **x** se compone de las variables de entrada, x_1, x_2, x_3 etc., que se utilizan para esa tarea.

Video explicativo

https://www.youtube.com/watch?v=JcI5E2Ng6r4&ab_channel=IntuitiveMachineLearning

https://www.youtube.com/watch?v=Ih3U8Rju5ck

Referencias

[1]https://es.wikipedia.org/wiki/Aprendizaje_basado_en_%C3%A1rboles_de_decisi%C3%B3n

[2]https://www.youtube.com/watch?v=JcI5E2Ng6r4&ab_channel=IntuitiveMachineLearning

[3]https://spark.apache.org/docs/2.4.7/mllib-decision-tree.html

[4https://spark.apache.org/docs/2.4.7/ml-classification-regression.html?fbclid=IwAR3QHShNZQ-gTK3XzVKacVE7NORmYZqX_74qqDw_Yr2lx1sA-nEJJcPh0Kw#decision-tree-classifier