

Árboles de decisión

Un **árbol de decisión** es un modelo de aprendizaje automático que se emplea en problemas de clasificación, donde siguiendo una estructura en forma de árbol se baja por las ramas en función de los valores de los atributos hasta llegar a un nodo hoja que contiene la clase.

El árbol de decisión se suele utilizar en problemas en los que se tienen una serie de atributos de los cuáles se quiere aprender y conocer más acerca de los valores y son útiles cuando se desea tener un modelo gráfico que clasifique la información.

Se puede considerar también una estructura en la que si nos movemos a través de él podemos llegar a un nodo que nos da una decisión final. Un ejemplo de este tipo de algoritmo es el caso el titanic que tenemos como entrada los datos de pasajeros del titanic y para cada pasajero tenemos el nombre, clase en la que viajó, cuanto pagó.. El objetivo es montar un árbol de decisión que nos diga para una determinada persona si sobrevivió o no.

Los **Arboles de Decision** son diagramas con construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

Los **Arboles de Decision** están compuestos por nodos interiores, nodos terminales y ramas que emanan de los nodos interiores. Cada nodo interior en el árbol contiene una prueba de un atributo, y cada rama representa un valor distinto del atributo. Siguiendo las ramas desde el nodo raíz hacia abajo, cada ruta finalmente termina en un nodo terminal creando una segmentación de los datos.

Los árboles de decisión son modelos ampliamente utilizados para las tareas de clasificación y regresión. En esencia, aprenden una jerarquía de preguntas if / else, lo que lleva a una decisión. Aprender un árbol de decisión significa aprender la secuencia de preguntas if / else que nos lleva a la respuesta verdadera más rápidamente.

A continuación muestro un ejemplo de un árbol de decisión:



La clave en árboles de decisión es encontrar un atributo cuyo valor nos permita separar una clase de todas las demás.

Dentro de las **ventajas** con que cuentan estos algoritmos están:

- Determinan aquellos atributos que aportan información relevante al sistema
- Establece jerarquías de importancia entre los atributos
- Rápido clasificando nuevos elementos
- No tiene gran coste computacional la creación del modelo
- El modelo puede ser modificable o pulido por un experto una vez creado
- Entendible e interpretable fácilmente
- El modelo se puede entender como un conjunto de reglas si , no, entonces

A grandes rasgos este clasificador es especialmente útil cuando los ejemplos a partir de los que se desea aprender se puedan representar mediante atributos y valores, así como cuando se desea tener un modelo gráfico que clasifique la información, sin embargo, no resultan adecuados cuando la estructura de los ejemplos es variable, es decir, cuando no conocemos todos los valores posibles o límites máximos y mínimos de los atributos. También presentan inconvenientes con información incompleta.

Existen diversas y variadas formas de construirlos, la tendencia general es utilizar una fórmula que nos indique qué atributo desplegar, como tal, las fórmulas más utilizadas son las siguientes:

Entropía

Describe la tendencia al orden en un sistema (también conocida como medida de incertidumbre). Básicamente determina en base a una situación, la probabilidad de que ocurra cada uno de los casos subsecuentes posibles. Va de 0 a 1, en donde 0 es el orden total y 1 el desorden total. Usualmente para los árboles de decisión suele utilizarse su forma binaria (al menos una forma de expresarla):

$$-\sum_{i=0}^{c-1} (p(i)) \log_2(p(i))$$

La entropía se define como el valor esperado de la información. Primero, necesitamos definir la información. Si está clasificando algo que puede tomar varios valores, la información para el símbolo x_i se define como

$$I(x_i) = -\log_2 p(x_i)$$

donde $p(x_i)$ es la probabilidad de elegir esta clase.

Para calcular entropía, se necesita el valor esperado de toda la información de todos los valores posibles de nuestra clase. Esto viene dado por la fórmula

$$H = -\sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

donde n es el número de clases.

Algoritmo ID3

http://en.wikipedia.org/wiki/ID3_algorithm

Algunos árboles de decisión hacen una división binaria de los datos.

Si dividimos un atributo y tiene cuatro valores posibles, entonces dividiremos los datos de cuatro maneras y crearemos cuatro ramas separadas. Seguiremos el algoritmo ID3, que nos dice cómo dividir los datos y cuándo dejar de dividirlo. También vamos a dividir en una y sólo una característica a la vez. Si nuestro conjunto de entrenamiento tiene 20 características, ¿cómo elegimos cuál usar primero? La respuesta a esta pregunta la tiene el término ganancia de información.

Ganancia de información

Escogemos dividir nuestro conjunto de datos de una manera que haga que nuestros datos no organizados estén más organizados. Hay múltiples maneras de hacer esto, y cada uno tiene sus propias ventajas y desventajas.

Una forma de organizar este desorden es medir la información. Utilizando la teoría de la información, puede medir la información antes y después de la división. La teoría de la información es una rama de la ciencia que se ocupa de cuantificar la información. El cambio en la información antes y después de la división se conoce como ganancia de información. Cuando se sabe cómo calcular la ganancia de información, puede dividir sus datos en todas las funciones para ver qué división le proporciona la mayor ganancia de información. La división con la mayor ganancia de información es la mejor opción. Antes de poder medir la mejor división y comenzar a dividir nuestros datos, debe saber cómo calcular la ganancia de información. La medida de la información de un conjunto se conoce como la entropía de Shannon, o simplemente la entropía para abreviar. Su nombre proviene del padre de la teoría de la información, Claude Shannon.

Selección de atributos

La mayoría de los algoritmos de aprendizaje automático están diseñados para aprender cuáles son los atributos más apropiados para usar para tomar sus decisiones. Por ejemplo, los métodos de árboles de decisión eligen el atributo más prometedor para dividirse en cada punto y deberían -en teoría- no seleccionar atributos irrelevantes o inútiles. Tener más características debería, en teoría, resultar en un poder más discriminatorio, nunca menos. "¿Cuál es la diferencia entre la teoría y la práctica?". En la práctica, agregar atributos irrelevantes o que distraen a un conjunto de datos a menudo confunde "los sistemas de aprendizaje automático".