

Concepto de clasificación

La clasificación es un sistema de agrupación, es asignar objetos a una de varias categorías, es decir clasificar es determinar la clase o grupo al cual pertenece un objeto, con el fin de ordenar y darle una identidad colectiva.

Es de suma importancia ya que nos sirve para organizar y ordenar. En nuestra vida diaria clasificamos las cosas que nos rodean, la mayoría del tiempo inconscientemente. El orden que obtenemos al agrupar lo que queremos conocer o las cosas que tenemos, hace que resulte más fácil nuestra tarea. Por lo cual es preciso decir que la clasificación esta presente en todos los ámbitos de nuestra vida, incluso nosotros mismo somos clasificados dependiendo la posición social que ocupemos.

Existen los modelos de clasificación descriptivos y predictivos. El descriptivo puede servir como una herramienta explicativa para distinguir entre objetos de diferentes clases, mientras que un modelo predictivo se puede usar para predecir el nivel de clase de registros desconocidos.

Un árbol de decisión es una forma de representar todos los sucesos que pueden surgir a partir de una decisión asumida en un cierto punto. Estos miden las probabilidades para ayudarnos a visualizar claramente el problema y tomar la decisión más acertada.

Entre los algoritmos más usados para construir un árbol de decisión está el algoritmo Hunt. En el algoritmo de Hunt, un árbol de decisión crece de forma recursiva al dividir los registros de entrenamiento en subconjuntos sucesivamente más puros. Sea D el conjunto de registros de entrenamiento asociados con los nodos t y $g: \{A_t, U_2, \dots, A_n\}$ sean las etiquetas de clase. La siguiente es una definición recursiva del algoritmo de Hunt.

- Paso 1: Si todos los registros en D_t pertenecen a la misma clase y_t , entonces t es un nodo hoja etiquetado como y_t .
- Paso 2: Si D_t contiene registros que pertenecen a más de una clase, se selecciona una condición de prueba de atributo para particionar los registros en subconjuntos más pequeños. Se crea un nodo secundario para cada resultado de la condición de prueba y los registros en D_t se distribuyen a los hijos según los resultados. El algoritmo se aplica recursivamente a cada nodo secundario

Es necesario tener en cuenta la medición del rendimiento del modelo de clasificación cuando se encuentra en la etapa de prueba, porque esto nos permitirá ver los errores del mismo. El margen de error que este esté presentando se puede utilizar para comparar el rendimiento relativo de diferentes clasificadores en el mismo dominio. Entre los métodos comúnmente utilizados para evaluar un clasificador están el aprendizaje supervisado y el aprendizaje no supervisado.