

Tarea 3

- Ignacio David Vázquez Pérez
- 218292866
- [\(2022B\) Sem. de Sol. de Problemas de Inteligencia Artificial II](#)
- [D05 \(2022B\)](#)

Realizar una investigación sobre clasificadores y reconocimiento de patrones distintos a los basados en perceptron.

Algunos ejemplos pueden ser:

- Deep Learning
- Extreme Learning Machines
- Clasificadores Lineales
- Clustering
- Lógica difusa

Se debe entregar en formato **pdf con una extensión de dos cuartillas**.

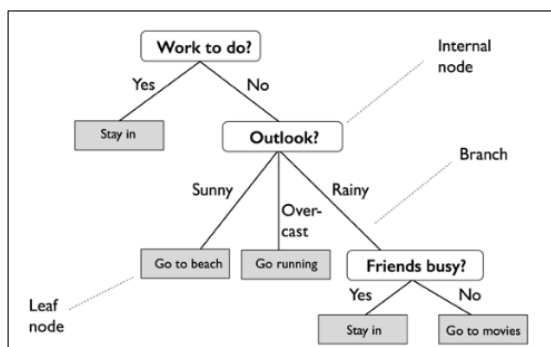
Se entrega de manera individual.

Introducción

Aprendizaje del árbol de decisión

Los clasificadores de árboles de decisión son modelos atractivos si nos preocupamos por la interpretabilidad. Como sugiere el nombre del árbol de decisiones, podemos pensar en este modelo como si desglosáramos nuestros datos al tomar una decisión basada en hacer una serie de preguntas.

Consideremos el siguiente ejemplo en el que usamos un árbol de decisiones para decidir sobre una actividad en un día en particular:



Según las características de nuestro conjunto de entrenamiento, el modelo de árbol de decisión aprende una serie de preguntas para inferir las etiquetas de clase de las muestras. Aunque la figura

anterior ilustra el concepto de un árbol de decisión basado en variables categóricas, el mismo concepto se aplica si nuestras características son números reales.

Usando el algoritmo de decisión, comenzamos en la raíz del árbol y dividimos los datos en la función que da como resultado la mayor ganancia de información (IG).

En un proceso iterativo, podemos repetir este procedimiento de división en cada nodo secundario hasta que las hojas estén puras. Esto significa que todas las muestras en cada nodo pertenecen a la misma clase. En la práctica, esto puede dar como resultado un árbol muy profundo con muchos nodos, lo que puede conducir fácilmente a un sobreajuste. Por lo tanto, normalmente queremos podar el árbol estableciendo un límite para la profundidad máxima del árbol.

Maximizar la ganancia de información

Para dividir los nodos en las características más informativas, necesitamos definir una función objetivo que queremos optimizar a través del algoritmo de aprendizaje del árbol. Aquí, nuestra función objetivo es maximizar la ganancia de información en cada división, que definimos de la siguiente manera:

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j)$$

Aquí, f es la función para realizar la división, D_p y D_j son el conjunto de datos del nodo padre y j -ésimo hijo, I es nuestra medida de impureza, N_p es el número total de muestras en el nodo padre y N_j es el número de muestras en el j -ésimo nodo secundario. Como podemos ver, la ganancia de información es simplemente la diferencia entre la impureza del padre nodo y la suma de las impurezas del nodo secundario: cuanto menor sea la impureza de los nodos secundarios, mayor será la ganancia de información.

Esto significa que cada nodo principal se divide en dos nodos secundarios, $D_{izquierdo}$ y $D_{derecho}$:

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right})$$

Ahora, las tres medidas de impureza o criterios de división que se usan comúnmente en los árboles de decisión binarios son la **impureza de Gini** (IG), la **entropía** (IH) y el **error de clasificación** (IE). la

- La definición de entropía para todas las clases no vacías $(p(i|t) \neq 0)$:

$$I_H(t) = - \sum_{i=1}^c p(i|t) \log_2 p(i|t)$$

Aquí, $p(i|t)$ es la proporción de las muestras que pertenecen a la clase i para un nodo particular t . Por lo tanto, la entropía es 0 si todas las muestras en un nodo pertenecen a la misma clase, y la entropía es máxima si tenemos una distribución de clases uniforme.

- La impureza de Gini puede entenderse como un criterio para minimizar la probabilidad de clasificación errónea:

$$I_G(t) = \sum_{i=1}^c p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^c p(i|t)^2$$

- Otra medida de impureza es el error de clasificación:

$$I_E = 1 - \max \{p(i|t)\}$$

Este es un criterio útil para la poda, pero no se recomienda para hacer crecer un árbol de decisión, ya que es menos sensible a los cambios en las probabilidades de clase de los nodos.

Los árboles de decisión pueden crear límites de decisión complejos dividiendo el espacio de características en rectángulos. Sin embargo, debemos tener cuidado ya que cuanto más profundo es el árbol de decisión, más complejo se vuelve el límite de decisión, lo que puede resultar fácilmente en un sobreajuste.

Lógica Difusa

Muchas tareas son simples para los humanos, pero crean un desafío continuo para las máquinas. Los ejemplos de tales sistemas incluyen caminar a través de un entorno desordenado, levantar objetos frágiles o estacionar un automóvil. La capacidad de los humanos para manejar datos vagos e imprecisos nos facilita esas tareas. Por lo tanto, si pretendemos replicar las acciones de control de un operador humano, debemos ser capaces de modelar las actividades del operador y no de la planta en sí.

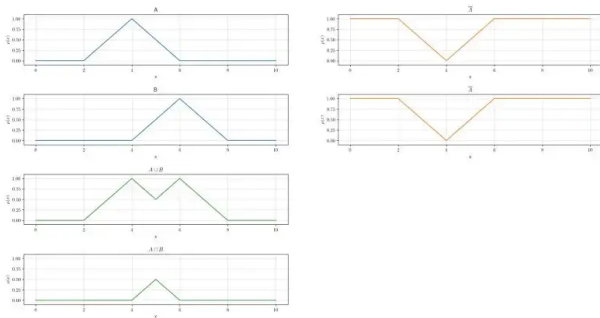
Nuestro modelo debe construirse de manera que sea capaz de manejar información vaga. Los sistemas basados en lógica difusa hacen precisamente eso; sobresalen donde los sistemas son particularmente complejos.

La lógica clásica se basa en el conjunto nítido, donde un grupo de objetos distintos se considera como una colección.

A diferencia de los conjuntos nítidos, un conjunto borroso permite la pertenencia parcial a un conjunto, que se define por un grado de pertenencia, denotado por μ , que puede tomar cualquier valor de 0 (el elemento no pertenece en absoluto al conjunto) a 1 (el elemento pertenece totalmente al conjunto).

Es evidente que si eliminamos todos los valores de pertenencia excepto el 0 y el 1, el conjunto borroso colapsará en un conjunto nítido. La función de pertenencia del conjunto es la relación entre los elementos del conjunto y su grado de pertenencia.

El método más común para calcular la unión de dos conjuntos borrosos es aplicando el operador máximo a los conjuntos. Existen otros métodos, incluido el uso del operador de producto en los dos conjuntos. De manera similar, el método más común para calcular la intersección de dos conjuntos borrosos es aplicar el operador mínimo en los conjuntos. El complemento de un conjunto borroso se calcula restando de 1 la función de pertenencia del conjunto.

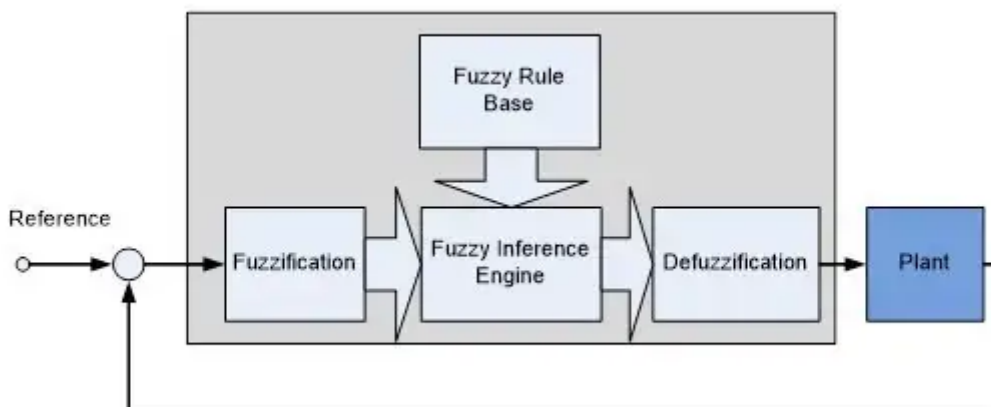


Sistemas de inferencia borrosa

Un sistema difuso es un depósito de conocimiento experto difuso que puede razonar los datos en términos vagos en lugar de una lógica booleana precisa. El conocimiento experto es una colección de funciones de pertenencia difusa y un conjunto de reglas difusas, conocido como base de reglas, que tiene la forma:

SI (se cumplen las condiciones) ENTONCES (se infieren las consecuencias)

La configuración básica de un sistema difuso se muestra a continuación:



Un sistema difuso típico se puede dividir en cuatro partes principales, a saber, un fuzzificador, una base de conocimientos, un motor de inferencia y un defuzzificador;

- El **fuzzificador** asigna una entrada nítida real a una función difusa, determinando así el "grado de pertenencia" de la entrada a un concepto vago. En varios controladores, los valores de las variables de entrada se asignan al rango de valores del universo de discurso correspondiente.
- La **base de conocimientos** comprende el conocimiento del dominio de la aplicación y los objetivos de control correspondientes. Se puede dividir en una base de datos de **definiciones** utilizadas

para expresar reglas de control lingüístico en el controlador y una **base de reglas** que describe el conocimiento que tienen los expertos del dominio. Intuitivamente, la base de conocimientos es el elemento central de un controlador difuso, ya que contendrá toda la información necesaria para realizar sus tareas de ejecución.

- El **motor de inferencia** proporciona la lógica de toma de decisiones del controlador. Deduce las acciones de control difuso empleando implicaciones difusas y reglas de inferencia difusas.
- El proceso **de Defuzzificación** convierte los valores de control borrosos en cantidades nítidas, es decir, vincula un solo punto a un conjunto borroso, dado que el punto pertenece al soporte del conjunto borroso.