

# Árboles de Decisión

Dra. Marcela Riccillo



# Tipos de Árboles de Decisión

- ▶ Árboles de Decisión para Clasificación - si la Var dependiente es cualitativa
- ▶ Árboles de Decisión para Regresión - si la Var dependiente es cuantitativa

Ejemplos:

- ▶ CART - ID3 - C4.5 - CHAID

# Posibles Criterios de Formación de árboles

+ Se seleccionan las variables según diferentes criterios

Entropía de Información (se busca bajar la Entropía)

$$\text{Entropía} \Rightarrow - \sum (p(x_i) * \log_2(p(x_i)))$$

Ganancia de Información

Índice de Gini

# Poda de árboles

Un árbol de decisión se puede desarrollar hasta ciertos parámetros, como mínima cantidad de elementos por nodo o profundidad del árbol (pre-poda)

También podría desarrollarse y luego ser podado (post-poda). Para eso se puede utilizar el parámetro de complejidad.

# Práctica (1/3)

Dada una base:

Mostrar un resumen de las características de los datos

¿Qué tipos de variables se pueden identificar?

¿Cuántas variables tiene la base? ¿Cuántos registros?

¿Cuál es la variable a predecir?

Particionar los datos en un conjunto de entrenamiento y uno de testeo

# Práctica (2/3)

Modelar el árbol. Analizar la estructura del árbol

Graficar el árbol

Considerar una de las hojas e indicar la regla para llegar desde la raíz del árbol hasta la misma

# Práctica (3/3)

Testear el árbol con el conjunto de testeo

Armar una matriz de confusión

¿Cuántos registros de cada clase quedaron bien y mal clasificados?

¿Cuál es la exactitud (accuracy) del modelo? ¿Cuál es la sensibilidad y la especificidad?