

Aprendizaje Automático

TP1: Método de Bayes

3 de marzo de 2021

1. Consideremos el siguiente vector de atributos binarios:

$(scones, cerveza, whisky, avena, futbol)$

El vector $x = (1, 0, 1, 1, 0)$ significa que se trata de una persona que le gustan los scones, no toma cerveza, le gusta el whisky y la avena pero no ve futbol. En el archivo `PreferenciasBritanicos.xls` se encuentran las preferencias de 6 personas inglesas y 7 personas escocesas.

- a) Implementar el clasificador ingenuo de Bayes.
 - b) Clasificar el ejemplo $x_1 = (1, 0, 1, 1, 0)$ determinando si corresponde a las preferencias de una persona inglesa o escocesa.
 - c) Clasificar el ejemplo $x_2 = (0, 1, 1, 0, 1)$, determinando si corresponde a las preferencias de una persona inglesa o escocesa.
2. Implementar un clasificador de texto utilizando el clasificador ingenuo de Bayes. Utilizar el conjunto de datos "Noticias Argentinas" para clasificar cada noticia según su tipo.
 - a) Utilizar al menos 4 categorías. Dividir el conjunto de textos disponible para utilizar una parte de los mismos como conjunto de entrenamiento y otro como conjunto test.
 - b) Construir la matriz de confusión.
 - c) Calcular las medidas de evaluación Accuracy, Precision, tasa de verdaderos positivos, tasa de falsos positivos y F_1 -score.
 - d) Calcular la curva ROC.
 3. El conjunto de datos `binary.csv` contiene información de la admisión de estudiantes a una universidad. Las variables son:
 - *admit*: (toma valores 0: no fue admitido, 1 fue admitido),
 - *GRE*: (Graduate Record Exam scores) variable numérica,
 - *GPA*: (grade point average) variable numérica,
 - *rank*: variable categórica que se refiere al prestigio de la escuela secundaria a la que el alumno asistió y toma valores $\{1, 2, 3, 4\}$.

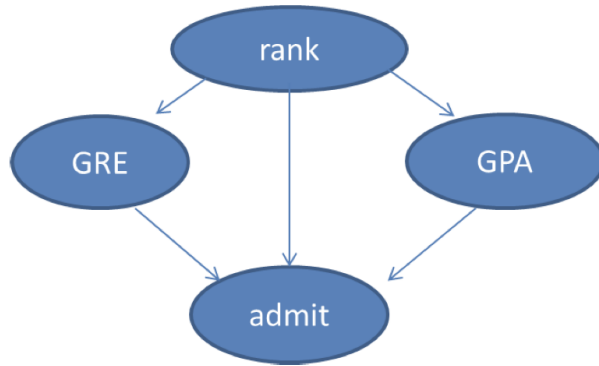


Figura 1: Relaciones entre las variables

Un investigador está interesado en averiguar cómo influyen estas variables en la admisión. Discretiza las variables GRE y GPA de la siguiente manera $GRE \in \{GRE \geq 500, GRE < 500\}$ y $GPA \in \{GPA \geq 3, GPA < 3\}$. Sabe que estas variables cumplen las relaciones presentadas en la Figura 1.

- a) Calcular la probabilidad de que una persona que proviene de una escuela con rango 1 no haya sido admitida en la universidad.
- b) Calcular la probabilidad de que una persona que fue a una escuela de rango 2, tenga $GRE = 450$ y $GPA = 3.5$ sea admitida en la universidad.
- c) En este ejercicio, ¿cuál es el proceso de aprendizaje?