Aprendizaje Automático
TP1: Método de Bayes

28 de agosto de 2024

1. Consideremos el siguiente vector de atributos binarios: (*scones, cerveza, whisky, avena, fútbol*)

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.
d. Detallar paso a paso cómo calcular las probabilidades para los incisos anteriores.

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.

   e. Utilizar al menos 4 categorías. Justifica la elección de las categorías utilizadas, por ejemplo en base a un análisis preliminar.

   f. Dividir el conjunto de textos disponible para utilizar una parte de los mismos como conjunto de entrenamiento y otro como conjunto test.

   g. Construir la matriz de confusión.

   h. Calcular las medidas de evaluación Accuracy, Precisión, tasa de verdaderos positivos, tasa de falsos positivos y F_1-score. Interpreta estos resultados en el contexto de las noticias clasificadas.

   i. Calcular la curva ROC y analizarla.

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 numerica
   - *GPA*: (grade point average) variable numerica,
   - *rank*: variable categórica que se refiere al prestigio de la escuela secundaria a la que el alumno asistió y toma valores $\{1, 2, 3, 4\}$

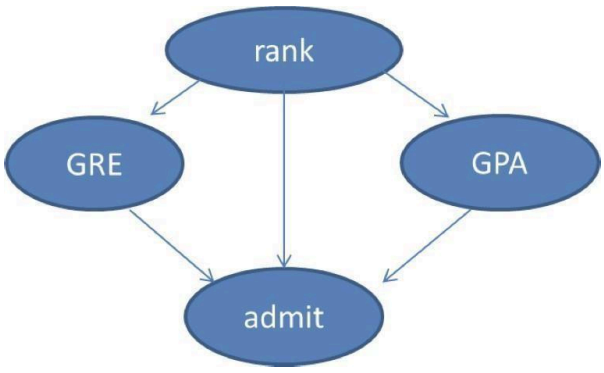

Figura 1: Relaciones entre las variables

Un investigador está interesado en averiguar cómo influyen estas variables en la admisi n. Discretizar las variables *GRE* y *GPA* de la siguiente manera $GRE \in \{GRE \geqslant 500, GRE < 500\}$ y $GPA \in \{GPA \geqslant 3, GPA < 3.\}$ 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. **Detallar paso a paso cómo calcular la probabilidad** y explicar todos los cálculos involucrados.

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. <u>En este ejercicio, ¿Cuál es el proceso de aprendizaje?</u>