

Elements of Machine Learning 2024

Tarea 04

03.mayo.2024

1. Construir un clasificador *Naïve Bayes* con los datos de la tabla `golf_data.csv`.
Elaborar un listado de todas las condiciones de clima en las cuáles sí conviene jugar golf, y un listado de todas las condiciones de clima en las cuales no conviene jugar.
2. Construya un modelo de clasificación (usted decide cuál algoritmo de clasificación aplicar) para clasificar un conjunto de datos de emails como *spam* o *no spam*.

La base de datos corresponde a estadísticas de 4601 emails, principalmente conteos de frecuencias relativas de ocurrencia de algunas palabras o símbolos particulares, y se encuentra en el archivo `spambase.zip`. Evaluar el desempeño del modelo, mediante métricas de clasificación (accuracy, F1-score, specificity, sensitivity), o mediante una matriz de confusión. No se olvide de separar sus datos en un conjunto de entrenamiento, uno de prueba.

Más información sobre este conjunto de datos puede hallarse en
<https://search.r-project.org/CRAN/refmans/kernlab/html/spam.html> y
<https://archive.ics.uci.edu/dataset/94/spambase>.

3. Los datos en el archivo `winequality-white.csv` corresponden a una evaluación de calidad sobre 4898 muestras de vinos, en función de indicadores y pruebas psicoquímicas.

El objetivo es modelar el valor de la variable $y = \text{quality}$, la última columna de la tabla, en función de las otras variables. Si lo considera conveniente, puede trabajar sobre una submuestra de los datos.

- (a) Haga un modelo de regresión lineal ordinaria (OLS) sobre esta base de datos, para estimar y .
- (b) Haga un modelo de regresión usando KNN, sobre esta base de datos, para estimar y .
- (c) Comparar ambos resultados obtenidos para un subconjunto de prueba (contrastar el valor estimado por ambos regresores, contra los valores reales de y) trabajo adecuado para ajustar los datos.
Calcule el error promedio MSE o MAE en ambos regresores. Discuta sus hallazgos, principalmente cuál de los dos métodos le parece más preciso.

Más información sobre este conjunto de datos puede hallarse en
<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>.
