

PRÁCTICA 8: REGRESIÓN LINEAL MÚLTIPLE Y LOGÍSTICA

1. Descargue el conjunto de datos “segment” del repositorio “openml.org”. Se proporciona el origen del dataset (<http://archive.ics.uci.edu/ml/datasets/image+segmentation>), aunque lo que hay en el repositorio “uci”, no es exactamente igual que lo hay en “openml”. A pesar de que está en diferentes formatos, quizás, el más conveniente sea el arff de WEKA que, para ser manejado en python junto sklearn, habrá que transformar a CSV, puesto que deberá ser leído por el módulo pandas correspondiente.
2. Con este fichero generado realice una clasificación mediante regresión lineal múltiple. Para ello, se destinará $\frac{2}{3}$ de los datos escogidos aleatoriamente de manera estratificada para aprendizaje y, el resto, para verificación.
3. Repetir el paso anterior empleando el método de validación cruzada con estratificación. Compare los resultados.
4. Repita los pasos 2 y 3 aplicando la Regresión logística. En ambos casos, solamente cuando emplee el método del resorte o Hold-Out, calcule la matriz de confusión.

NOTA: Este trabajo se hará en python usando el entorno jupyter, en donde, aparte del código, deberán aparecer tantas cajas como estime oportuno para diferenciar claramente los puntos anteriores, así como cuantas divisiones estime oportunas para hacer un mejor seguimiento y explicación de los pasos dados para la resolución de esta práctica. Por tanto, sólo será preciso subir el fichero correspondiente fichero ‘ipynb’ generado por jupyter.