



Análisis de Datos



Jesús García Herrero

Práctica 1 WEKA. Clasificación con herramienta de análisis WEKA

Introducción a Weka.

En este ejercicio se trata de reproducir con la herramienta weka algunos de los resultados obtenidos con los ejemplos numéricos de la práctica anterior, analizados mediante clasificadores implementados en Excel. Se tienen los ficheros “datosAleatorios2D.arff”, que representa datos generados con variables aleatorias normales bidimensionales (150 muestras), y el fichero “titanic.arff”.

- a) Entrar en weka y cargar cada fichero.
- b) Visualizar los ejemplos en el caso numérico, representando los atributos “X” e “Y”, y la clase mediante código de colores.
- c) Analizar el comportamiento (error sobre el propio conjunto de entrenamiento) de un clasificador bayesiano (NaiveBayes) y de un clasificador mediante regresión lineal (ClassifierViaRegression). Comprobar si los resultados son consistentes con los ya conocidos.
- d) A continuación analizar el comportamiento de diferentes clasificadores: *Zero-R*, *One-R*, *ID3*, *C4.5 (J.48)*. Obsérvese que el clasificador ID3 únicamente es aplicable con atributos simbólicos.
- e) Analizar el efecto de realizar clasificaciones con costes asimétricos (CostSensitiveClassifier) sobre los errores globales y particulares (matriz de confusión).
- f) Comparar el resultado obtenido evaluando sobre el conjunto de entrenamiento con una evaluación mediante validación cruzada. Realizar evaluaciones sobre conjuntos de evaluación independientes, dividiendo el fichero de datos original en una parte de entrenamiento y otra de evaluación. Obsérvese qué ocurre si una clase no está bien representada.