

INTELIGENCIA ARTIFICIAL

PRÁCTICA 11



GRUPO 3:
Enrique Ballesteros Horcajo
Ignacio Iker Prado Rujas

7 de Mayo de 2015

Índice

1.	Introducción	1
2.	Apartado 1	1
3.	Apartado 2	1

1. Introducción

En esta práctica vamos a trabajar con el entorno que nos proporciona WEKA¹, aplicando herramientas de aprendizaje automático a conjuntos de datos dados por los ficheros `diabetes.arff` y `glass.arff`.

2. Apartado 1

Tabla 4x4. Ordenar con itemize o lo que sea para que quede elegante.

Analiza el archivo citado y contesta las siguientes preguntas: ¿Cuántas instancias hay? ¿Cuántos atributos se utilizan? ¿Por qué atributo queremos aprender a clasificar? - Ejecuta J48 (versión WEKA de C4.5). Utiliza para la validación el “training set”. Incluye en la memoria los resultados obtenidos y la representación gráfica del árbol correspondiente. ¿Cuántas instancias han sido mal clasificadas? ¿Cuántos nodos terminales hay en el árbol? ¿Por qué atributo se clasifica en el primer nivel del árbol? - Vuelve a ejecutar J48 primero con “cross-validation” y después con “percentage split” (con un valor del 66 %) y observa las diferencias. Haz una tabla con el porcentaje de instancias bien clasificadas en cada uno de los tres métodos de validación. Incluye en esa tabla la precisión y el recall para cada uno de los tres métodos. Comenta los resultados obtenidos. ¿Cuál de las tres validaciones te parece más fiable? ¿Por qué? ¿Cuántos falsos negativos (FN) y falsos positivos (FP) se han obtenido con el “percentage split”?

3. Apartado 2

Analiza el dataset. ¿En cuántas clases se clasifican los ejemplares? ¿Cuántos atributos se utilizan? - Aplica el clustering jerárquico a este conjunto de datos para 7 clusters y comenta los resultados obtenidos. ¿Qué clases han quedado mejor clasificadas? - Discretiza los atributos utilizando un filtro supervisado. ¿Qué horquillas se han generado para los valores del Calcio? Vuelve a ejecutar el clustering jerárquico y compara los resultados con los anteriores. ¿Qué clases han quedado mejor clasificadas? - Con los atributos en su estado original, utiliza el algoritmo de clustering EM (Expectation Maximization) y compara con los resultados anteriores. ¿Qué clases han quedado mejor clasificadas? ¿Qué método de los tres anteriores te parece más útil para este ejemplo? ¿Por qué? ¿Podríamos haber utilizado un algoritmo de clasificación como J48? ¿Qué habríamos obtenido con los algoritmos de clustering si no tuviéramos el atributo de clase? ¿En qué casos sólo podríamos utilizar clustering?

¹Página web de WEKA [aquí](#).

Bibliografía

- [1] Russell, S.; Norvig, P,
Artificial Intelligence, a modern aproach.
New Jersey: Pearson, 2010.
- [2] Apuntes y transparencias de Inteligencia Artificial,
Doble Grado Matemáticas - Ing. Informática, U.C.M., 2014-2015.