



INGENIERÍA DEL CONOCIMIENTO

Gonzalo Pajares

Práctica 3: Métodos de clasificación

1.- Preliminares

Se trata de implementar al menos un método de clasificación de entre los cuatro siguientes:

- Agrupamiento borroso (K-medias)
- Bayes
- Algoritmo de Lloyd
- Mapas Auto-organizativos (SOM)

NOTA: en el apéndice más abajo se indican los parámetros que requiere cada algoritmo.

2.- Planteamiento

- Dado el fichero *Iris2Clases.txt* con cuatro atributos numéricos pertenecientes a dos clases: *Iris-setosa* e *Iris-versicolor*. Estos datos se encuentran en el siguiente repositorio, ampliamente conocido a nivel mundial en temas relacionados con problemas de clasificación **UCI Machine Learning Repository** <http://archive.ics.uci.edu/ml/datasets.html>
- Leer el fichero de texto y agrupar los datos como pertenecientes a cada una de sus clases correspondientes.
- Implementar uno de entre los métodos propuestos en el punto uno.
- Proporcionarle al menos tres ejemplos, como los suministrados en los ejemplos de los ficheros *TestIris01.txt*, *TestIris02.txt* y *TestIris03.txt* y obtener el resultado de la clasificación según el método seleccionado.

3.- Posibilidades de ampliación y mejora de la práctica

- Implementar los cuatro métodos propuestos.
- Comprobar el resultado de la clasificación para los tres ejemplos (*TestIris01.txt*, *TestIris02.txt* y *TestIris03.txt*) en cada uno de los métodos.

4.- Entregables

Se entregará una memoria de la práctica consistente en:

- Describir el sistema mediante una tarea de Clasificación de CommonKADS. Similar al caso de la práctica dos, al tratarse de la misma tarea con la única variación de los métodos.
- Describir los detalles de implementación propios, tales como: lenguaje utilizado, procedimiento seguido para su implementación, ampliaciones realizadas y todos aquellos elementos considerados de interés.
- Código ejecutable
- Breve manual de usuario descriptivo para poder ejecutar la práctica.

APÉNDICE: Parámetros de los algoritmos

Inicialización de los centros de las clases: se puede realizar de forma aleatoria, los siguientes valores son apropiados

$$\mathbf{v} = \begin{bmatrix} 4.6 & 3.0 & 4.0 & 0.0 \\ 6.8 & 3.4 & 4.6 & 0.7 \end{bmatrix}$$

1.- Agrupamiento borroso (K-medias):

Tolerancia $\varepsilon = 0.01$

Peso exponencial $b = 2$;

2.- Bayes: no necesita inicialización

3.- Algoritmo de Lloyd:

Tolerancia $\varepsilon = 10^{-10}$

Número máximo iteraciones $k_{max} = 10$

Razón de aprendizaje (constante) = 0.1

4.- Mapas Auto-Organizativos (SOM):

Tolerancia $\varepsilon = 10^{-6}$

Número máximo iteraciones $k_{max} = 1000$

Razón de aprendizaje (constante) = 0.1

$\alpha_{inicial} = 0.1$; $\alpha_{final} = 0.01$

Consideramos que un centro se actualiza cuando $K_{\alpha(k)} > T$, siendo $T = 10^{-5}$