
Métodos de clasificación

Pablo Arranz Roper

Introducción

Dada la base matemática de esta práctica, he decidido implementarla en Octave. El código implementado creará un sistema de clasificación, que dados unos datos de entrada decidirá a que clase pertenece ese ejemplo. Se implementarán varios métodos de clasificación:

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

Para poder leer el fichero desde octave se ha transformado la última columna del fichero en 0 en el caso de Iris-setosa y 1 en el caso de Iris-versicolor. Para esto hemos usado las siguientes expresiones regulares.

```
sed -i -E 's/^(.*)Iris-setosa\10/g' Iris2Clases.txt
```

```
sed -i -E 's/^(.*)Iris-versicolor\11/g' Iris2Clases.txt
```

Para desarrollar los diferentes algoritmos he creado una función principal para cada uno y diferentes funciones auxiliares que eran útiles para hacer el código más modular.

Las funciones auxiliares creadas finalmente han sido:

- `gradopertenencia` para calcular el grado de pertenencia de un ejemplo a una clase determinada.
- `norma` para calcular la norma vectorial entre dos puntos.
- `vecindad` para calcular la región de vecindad en el algoritmo SOM.

Las funciones principales (`kmedias`, `bayes`, `lloyd`, `som`) devuelven cada una, respectivamente:

- Los centros y los grados de pertenencia (De los casos de test).
- Las matrices de covarianza.
- Los centros (en los dos últimos algoritmos).

Además de esto imprimen por pantalla para los 3 casos de test la clase a la que pertenecen, a la que deberían pertenecer y si es correcta su clasificación.

Manual de uso

Para poder iniciar la aplicación hay que tener instalado en el ordenador Octave. Tendremos que ir a la carpeta donde se encuentran los archivos y ejecutar desde la consola la orden `octave` para entrar en el entorno de octave. Una vez dentro las instrucciones a ejecutar son las siguientes para cada algoritmo.

K-medias

`[v, U] = kmedias()` devolverá el siguiente resultado:

```
[octave:31> [v, U] = kmedias()
is: Iris-setosa ; classified as: Iris-setosa ;   right
is: Iris-versicolor ; classified as: Iris-versicolor ;   right
is: Iris-setosa ; classified as: Iris-setosa ;   right
v =
    5.00455    3.40234    1.48737    0.25301
    5.97539    2.79314    4.30551    1.33900

U =
    9.9736e-01    7.2934e-02    9.9971e-01
    2.6364e-03    9.2707e-01    2.8556e-04
```

`kmedias()` ; tan solo mostrará por pantalla las 3 primeras líneas.

Bayes

`[C1, C2] = bayes()` devolverá el siguiente resultado:

```
[octave:32> [C1, C2] = bayes()
is: Iris-setosa ; classified as: Iris-setosa ;   right
is: Iris-versicolor ; classified as: Iris-versicolor ;   right
is: Iris-setosa ; classified as: Iris-setosa ;   right
C1 =
    0.1217640    0.0982920    0.0158160    0.0103360
    0.0982920    0.1422760    0.0114480    0.0112080
    0.0158160    0.0114480    0.0295040    0.0055840
    0.0103360    0.0112080    0.0055840    0.0112640

C2 =
    0.261104    0.083480    0.179240    0.054664
    0.083480    0.096500    0.081000    0.040380
    0.179240    0.081000    0.216400    0.071640
    0.054664    0.040380    0.071640    0.038324
```

`bayes()` ; tan solo mostrará por pantalla las 3 primeras líneas.

Lloyd

`v = lloyd()` devolverá el siguiente resultado:

```
[octave:35> v = lloyd()
is: Iris-setosa ; classified as: Iris-setosa ;   right
is: Iris-versicolor ; classified as: Iris-versicolor ;   right
is: Iris-setosa ; classified as: Iris-setosa ;   right
v =
    4.95783    3.37730    1.46849    0.25133
    5.74403    2.75161    4.09664    1.27468
```

`lloyd()` ; tan solo mostrará por pantalla las 3 primeras líneas.

SOM(Self-Organizative Maps)

`v = som()` devolverá el siguiente resultado:

```
[octave:36> v = som()
is: Iris-setosa ; classified as: Iris-setosa ;   right
is: Iris-versicolor ; classified as: Iris-versicolor ;   right
is: Iris-setosa ; classified as: Iris-setosa ;   right
v =
    4.60000    3.00000    4.00000    0.00000
    6.80000    3.40000    4.60000    0.70000
```

`som()` ; tan solo mostrará por pantalla las 3 primeras líneas.