

Ingeniería del Conocimiento

Métodos de clasificación.



Índice:

- Introducción.
- Algoritmos.
- Implementación.
- Funcionamiento de la app.

Introducción:

Este documento detalla la memoria de la práctica 3 de Ingeniería del Conocimiento, la implementación de los algoritmos de clasificación, K-Medias, Bayes y LLoyd.

Estos algoritmos nos permiten clasificar elementos en clases, dadas las características de la clase y las del elemento.

Ejemplo del archivo de entrada:

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
```

Ejemplo del elemento a clasificar:

```
5.1,3.5,1.4,0.2
```

Algoritmos:

K-Medias: es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano. Es un método utilizado en minería de datos.

El término "k-means" fue utilizado por primera vez por James MacQueen en 1967, aunque la idea se remonta a Hugo Steinhaus en 1957. El algoritmo estándar fue propuesto por primera vez por Stuart Lloyd en 1957 como una técnica para modulación por impulsos codificados, aunque no se publicó fuera de los laboratorios Bell hasta 1982. En 1965, E. W. Forgy publicó esencialmente el mismo método, por lo que a veces también se le nombra como Lloyd-Forgy. Una versión más eficiente fue propuesta y publicada en Fortran por Hartigan y Wong en 1975/1979.

Complejidad:

$O(n^{dk+1} \log n)$, donde n es el número de entidades a particionar

Bayes: El teorema de Bayes, en la teoría de la probabilidad, es una proposición planteada por el filósofo inglés Thomas Bayes (1702-1761) en 1763, que expresa la [probabilidad condicional](#) de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la [distribución de probabilidad marginal](#) de sólo A .

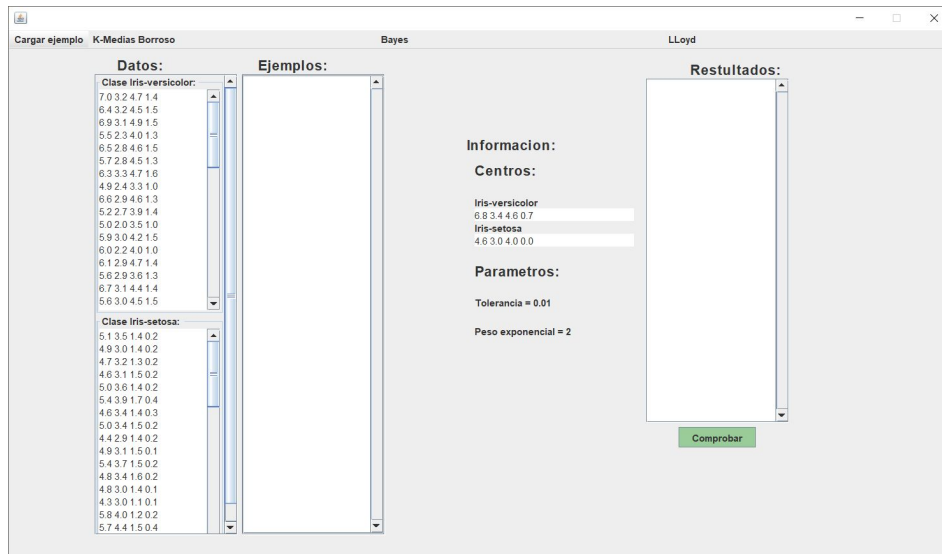
El teorema de Bayes es de enorme relevancia ya que vincula la probabilidad de A dado B con la probabilidad de B dado A . Es decir, por ejemplo, que sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber, la probabilidad de tener gripe si se tiene un dolor de cabeza.

Lloyd: El algoritmo de Lloyd es una técnica de agrupamiento que permite unificar n objetos en k clases. Al agrupar los objetos en distintos grupos o clases obtenemos un método de predicción que podemos aplicar tras encontrar un nuevo elemento. Así mismo podemos establecerlo en una de las clases conocidas, esperando que sus características sean similares a las del resto de elementos de dicha clase.

El algoritmo de Lloyd es un método de aprendizaje no supervisado debido a que nos permite obtener información sobre los objetos sin que aportemos de forma explícita un conocimiento previo sobre los mismos.

Implementación:

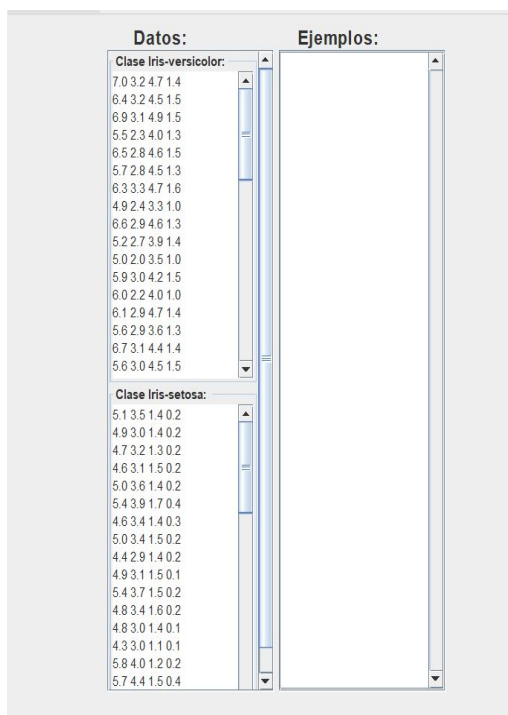
El lenguaje utilizado para el desarrollo de la práctica es Java, la interfaz gráfica consta de una sola ventana dividida en tres partes, un menú, un panel fijo que mantiene los datos introducidos y los ejemplos, y otro panel en el que se mostraran los diferentes algoritmos.



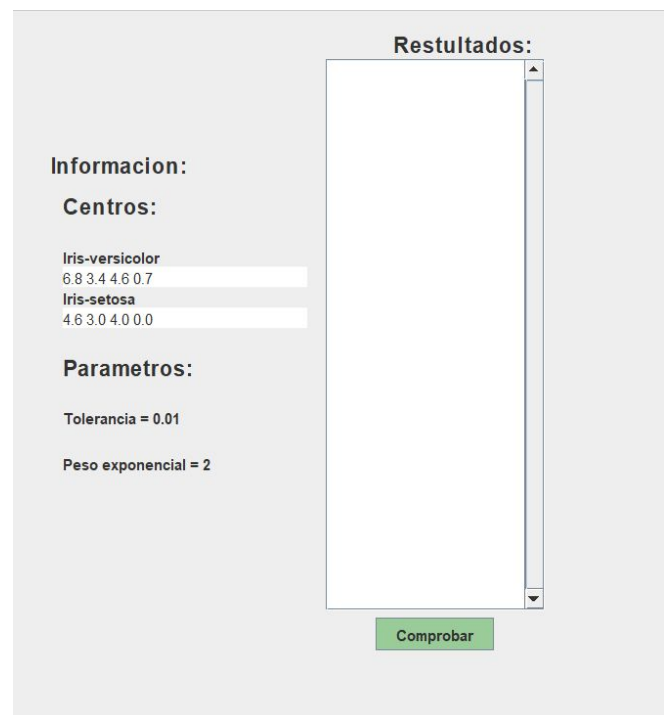
Menú



Panel de datos y ejemplos.



Panel de algoritmo.



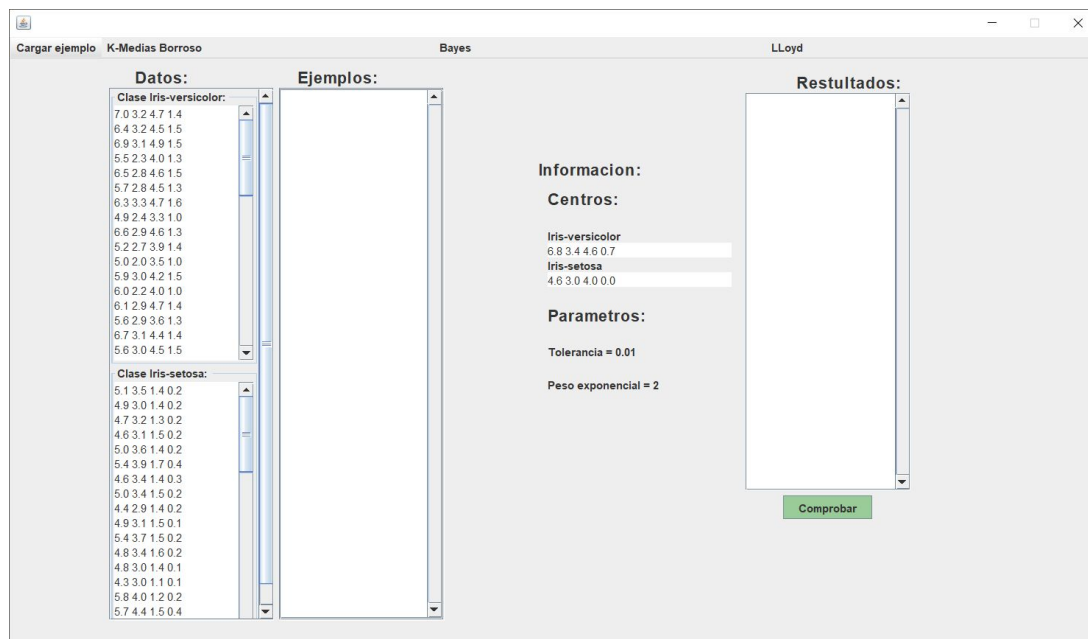
La aplicación busca, a través de diferentes estructuras dinámicas tales como HashMap, utilizada en la lectura u ArrayList<double>, para un posible incremento de clases, aceptando de esta manera una entrada con n clases, y n elementos de cada clase. Los centroides de las diferentes clases se han sacado de la aplicación, y guardados en documento de texto, pensando en esa posible ampliación de clases y modificación de los elementos.

Los algoritmos están separados y son independientes, reciben matrices bidimensionales, para los cálculos, estas matrices son creadas previamente, “parseadas” desde ArrayList<double[]> para aceptar n elemento de cada clase, el array de double, proviene a su vez de otro ArrayList<double> para aceptar un número n de características de clase.

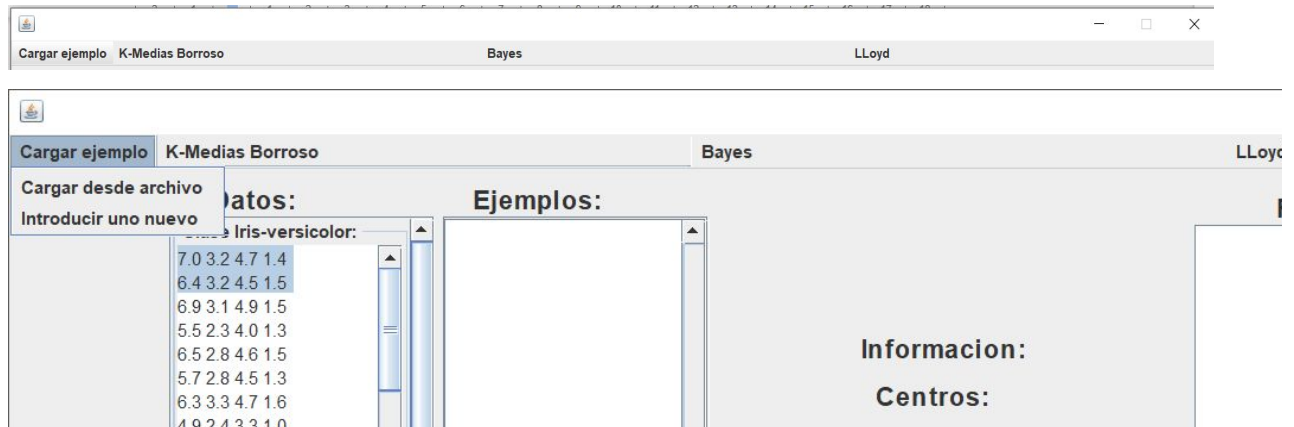
Los archivos necesarios para la ejecución de la aplicación son dos: Iris2Clases.txt, Centros.txt, los archivos de test pueden ser leídos dentro la aplicación.

Funcionamiento de la aplicación:

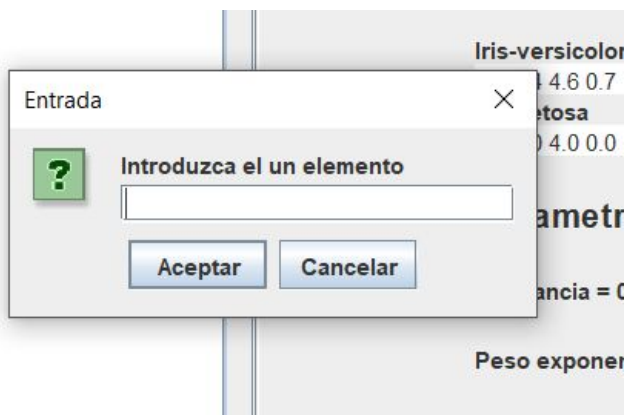
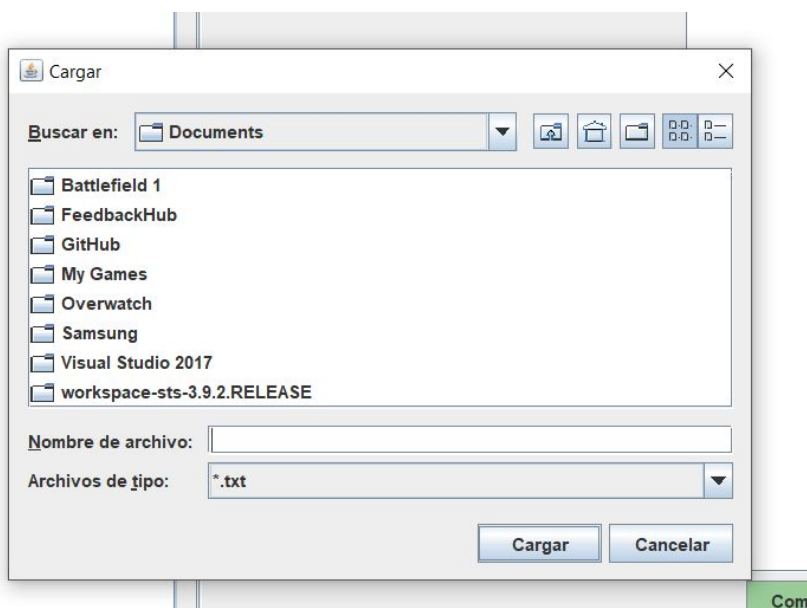
Una vez ejecutada la aplicación, si la carga de archivos es correcta se podrá ver la aplicación, tal cual la imagen:



En el menú superior podremos elegir, cualquiera de los tres algoritmos implementados, K-Medias, que es el algoritmo por defecto, Bayes y LLoyd. Además el primer botón nos permitirá la carga de ejemplos, para su posterior análisis.



Podremos cargar ejemplos de dos maneras diferentes, mediante un archivo de texto, como los entregados en la práctica, o utilizando un cuadro de texto:



Una vez se hayan introducido todos los datos de ejemplo que se deseen, se podrá, pulsar el botón de comprobar para obtener el resultado.

Cargar ejemplo

K-Medias Borroso

Bayes

LLoyd

Datos:

Clase Iris-versicolor:

7.0 3.2 4.7 1.4
6.4 3.2 4.5 1.5
6.9 3.1 4.9 1.5
5.5 2.3 4.0 1.3
6.5 2.8 4.6 1.5
5.7 2.8 4.5 1.3
6.3 3.3 4.7 1.6
4.9 2.4 3.3 1.0
6.6 2.9 4.6 1.3
5.2 2.7 3.9 1.4
5.0 2.0 3.5 1.0
5.9 3.0 4.2 1.5
6.0 2.2 4.0 1.0
6.1 2.9 4.7 1.4
5.6 2.9 3.6 1.3
6.7 3.1 4.4 1.4
5.6 3.0 4.5 1.5

Clase Iris-setosa:

5.1 3.5 1.4 0.2
4.9 3.0 1.4 0.2
4.7 3.2 1.3 0.2
4.6 3.1 1.5 0.2
5.0 3.6 1.4 0.2
5.4 3.9 1.7 0.4
4.6 3.4 1.4 0.3
5.0 3.4 1.5 0.2
4.4 2.9 1.4 0.2
4.9 3.1 1.5 0.1
5.4 3.7 1.5 0.2
4.8 3.4 1.6 0.2
4.8 3.0 1.4 0.1
4.3 3.0 1.1 0.1
5.8 4.0 1.2 0.2
5.7 4.4 1.5 0.4

Ejemplos:

6.9 3.1 4.9 1.5
6.9 3.1 4.9 1.5
5.1 3.5 1.4 0.2

Información:

Centros:

Iris-versicolor

6.8 3.4 4.6 0.7

Iris-setosa

4.6 3.0 4.0 0.0

Parametros:

Tolerancia = 0.01

Peso exponencial = 2

Resultados:

Iris-versicolor
Iris-versicolor
Iris-setosa

Comprobar

Presentación de los diferentes algoritmos:
K-Medias:

Cargar ejemplo

K-Medias Borroso

Bayes

Lloyd

Datos:

Clase Iris-versicolor:

7.0324714
6.4324515
6.9314915
5.5234013
6.5284615
5.7284513
6.3334716
4.9243310
6.6294613
5.2273914
5.0203510
5.9304215
6.0224010
6.1294714
5.6293613
6.7314414
5.6304515

Clase Iris-setosa:

5.1351402
4.9301402
4.7321302
4.6311502
5.0361402
5.4391704
4.6341403
5.0341502
4.4291402
4.9311501
5.4371502
4.8341602
4.8301401
4.3301101
5.8401202
5.7441504

Ejemplos:

6.9314915
6.9314915
5.1351402

Resultados:

Iris-versicolor
Iris-versicolor
Iris-setosa

Informacion:

Centros:

Iris-versicolor

6.8344607

Iris-setosa

4.6304000

Parametros:

Tolerancia = 0.01

Peso exponencial = 2

Comprobar

Bayes:

Cargar ejemplo

K-Medias Borroso

Bayes

Lloyd

Datos:

Clase Iris-versicolor:

7.0324714
6.4324515
6.9314915
5.5234013
6.5284615
5.7284513
6.3334716
4.9243310
6.6294613
5.2273914
5.0203510
5.9304215
6.0224010
6.1294714
5.6293613
6.7314414
5.6304515

Clase Iris-setosa:

5.1351402
4.9301402
4.7321302
4.6311502
5.0361402
5.4391704
4.6341403
5.0341502
4.4291402
4.9311501
5.4371502
4.8341602
4.8301401
4.3301101
5.8401202
5.7441504

Ejemplos:

6.9314915
6.9314915
5.1351402

Resultados:

Iris-versicolor
Iris-versicolor
Iris-setosa

Comprobar

Lloyd:

Cargar ejemplo

K-Medias Borroso

Bayes

Lloyd

Datos:

Clase Iris-versicolor:

7.0324714
6.4324515
6.9314915
5.5234013
6.5284615
5.7284513
6.3334716
4.9243310
6.6294613
5.2273914
5.0203510
5.9304215
6.0224010
6.1294714
5.6293613
6.7314414
5.6304515

Clase Iris-setosa:

5.1351402
4.9301402
4.7321302
4.6311502
5.0361402
5.4391704
4.6341403
5.0341502
4.4291402
4.9311501
5.4371502
4.8341602
4.8301401
4.3301101
5.8401202
5.7441504

Ejemplos:

6.9314915
6.9314915
5.1351402

Resultados:

Iris-versicolor
Iris-versicolor
Iris-setosa

Informacion:

Centros:

Iris-versicolor

6.8344607

Iris-setosa

4.6304000

Parametros:

Tolerancia = 0.000000001

Máximo de iteraciones = 10

Razón de aprendizaje = 0.1

Comprobar