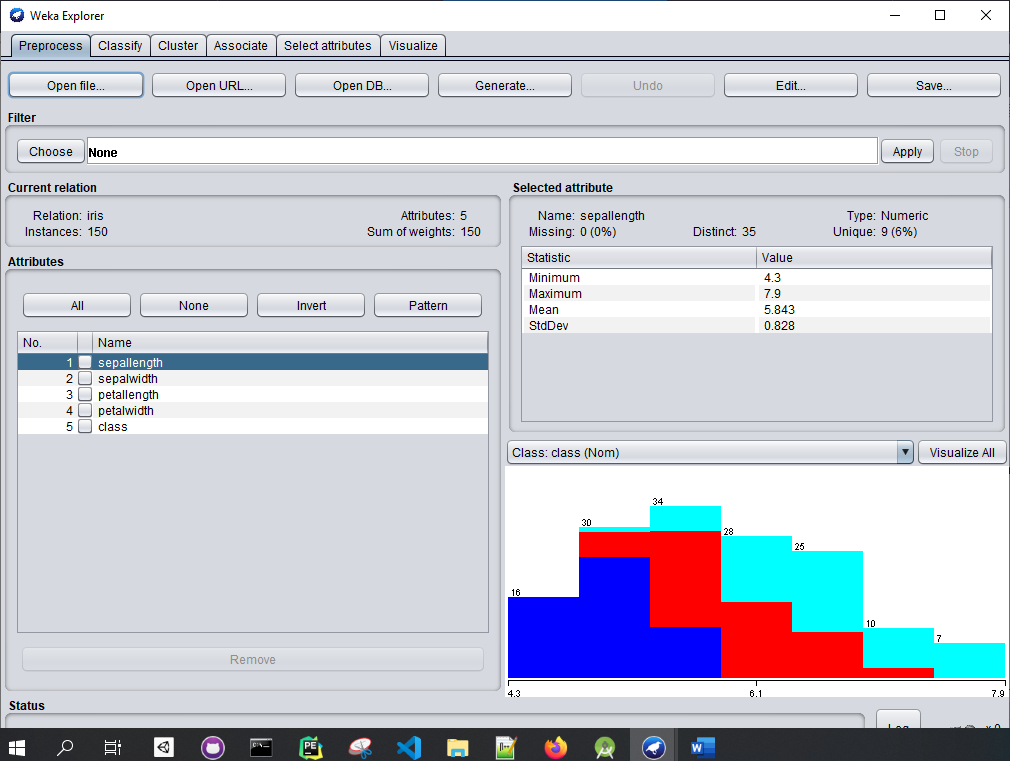
**Practica 4: Clustering**

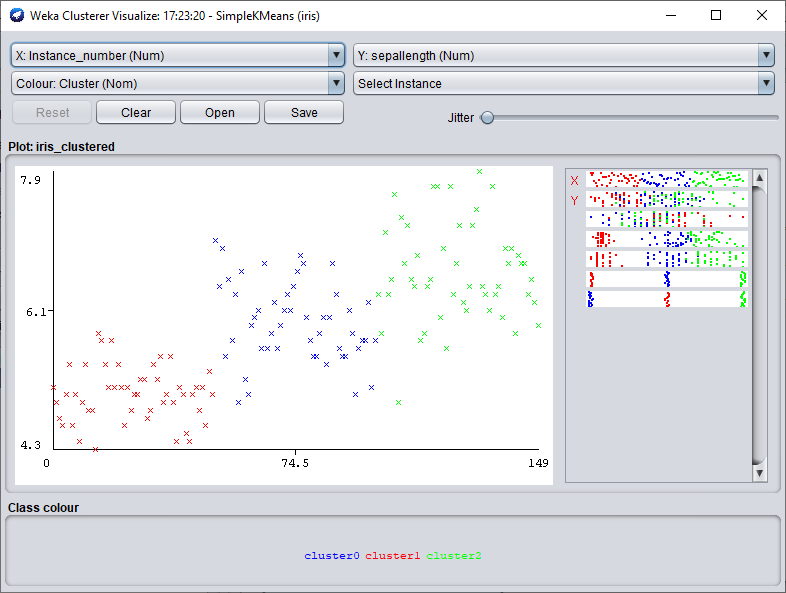
Práctica 4 : [Clustering](https://educacionvirtual.epn.edu.ec/mod/resource/view.php?id=380043)

Link de apoyo

https://www.youtube.com/watch?v=HCA0Z9kL7Hg

* Utilizando el archivo iris.arff generar 3 clusters usando el algoritmo Kmeans.

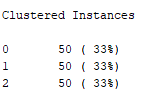




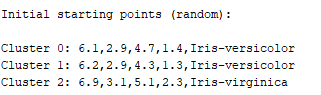
* Escoger como medida de distancia la distancia euclidiana



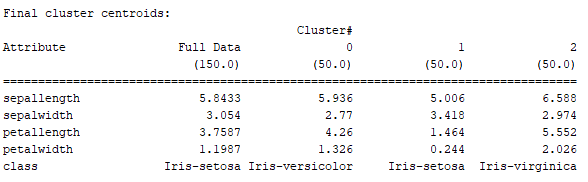
* Indicar el número de instancias en cada cluster



* Los centroides iniciales



* Los centroides finales



* La suma de los cuadrados de los errores



Ejercicio 1 Capítulo 19, Principlesof Datamining, Bramer

1. Usando el método que se muestra en la Sección 19.2, agrupe los siguientes datos en

tres grupos, utilizando el método k-means.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Punto | **x** | **y** | Cluster final | Cluster final | Cluster final |
| 1 | 10,90 | 12,60 | 2 | 2 | 2 |
| 2 | 2,30 | 8,40 | 1 | 1 | 1 |
| 3 | 8,40 | 12,60 | 1 | 2 | 2 |
| 4 | 12,10 | 16,20 | 3 | 3 | 3 |
| 5 | 7,30 | 8,90 | 1 | 2 | 2 |
| 6 | 23,40 | 11,30 | 3 | 3 | 3 |
| 7 | 19,70 | 18,50 | 3 | 3 | 3 |
| 8 | 17,10 | 17,20 | 3 | 3 | 3 |
| 9 | 3,20 | 3,40 | 1 | 1 | 1 |
| 10 | 1,30 | 22,80 | 1 | 1 | 1 |
| 11 | 2,40 | 6,90 | 1 | 1 | 1 |
| 12 | 2,40 | 7,10 | 1 | 1 | 1 |
| 13 | 3,10 | 8,30 | 1 | 1 | 1 |
| 14 | 2,90 | 6,90 | 1 | 1 | 1 |
| 15 | 11,20 | 4,40 | 2 | 2 | 2 |
| 16 | 8,30 | 8,70 | 2 | 2 | 2 |

Centroides de las 3 iteraciones







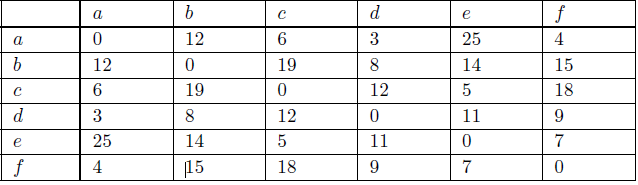
**Iteración 1**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Distancia 1 | Distancia 2 | Distancia 3 | Cluster final |  |  |  |  |
| 10,23 | 8,21 | 10,59 | 2 |  |  |  |  |
| 1,50 | 9,76 | 20,12 | 1 |  |  |  |  |
| 8,28 | 8,66 | 12,75 | 1 |  |  |  |  |
| 13,44 | 11,83 | 7,94 | 3 |  |  |  |  |
| 5,29 | 5,95 | 15,68 | 1 |  |  |  |  |
| 21,46 | 14,02 | 8,10 | 3 |  |  |  |  |
| 20,83 | 16,46 | 0,00 | 3 |  |  |  |  |
| 17,95 | 14,09 | 2,91 | 3 |  |  |  |  |
| 3,59 | 8,06 | 22,37 | 1 |  |  |  |  |
| 15,94 | 20,89 | 18,90 | 1 |  |  |  |  |
| 0,00 | 9,15 | 20,83 | 1 |  |  |  | 2 |
| 0,20 | 9,20 | 20,72 | 1 |  |  |  |  |
| 1,57 | 8,99 | 19,48 | 1 |  |  |  |  |
| 0,50 | 8,67 | 20,42 | 1 |  |  |  |  |
| 9,15 | 0,00 | 16,46 | 2 |  |  |  |  |
| 6,17 | 5,19 | 15,03 | 2 |  |  |  |  |
| **Iteración 2** |  |  |  |  |  |  |  |
| Distancia 1 | Distancia 2 | Distancia 3 |  |  |  |  |  |
| 7,847819545 | 4,105551797 | 7,856247514 | 2 | Igual |  |  |  |
| 1,766806424 | 7,835106182 | 17,4244261 | 1 | Igual |  |  |  |
| 5,642541236 | 4,390013921 | 10,19046736 | 2 | Diferente |  |  |  |
| 10,75863707 | 7,882610961 | 5,988374153 | 3 | Igual |  |  |  |
| 3,646070098 | 2,852873795 | 12,79494529 | 2 | Diferente |  |  |  |
| 19,78409699 | 13,54531489 | 6,971773447 | 3 | Igual |  |  |  |
| 18,36846466 | 13,79101962 | 3,151289419 | 3 | Igual |  |  |  |
| 15,46585646 | 11,09364182 | 1,706055392 | 3 | Igual |  |  |  |
| 6,098309825 | 8,646707787 | 19,36557835 | 1 | Igual |  |  |  |
| 13,53667629 | 16,75158367 | 18,17692562 | 1 | Igual |  |  |  |
| 2,887029316 | 7,910892631 | 18,02541608 | 1 | Igual |  |  |  |
| 2,70994966 | 7,871185143 | 17,92751028 | 1 | Igual |  |  |  |
| 1,321801987 | 7,03838681 | 16,74815288 | 1 | Igual |  |  |  |
| 2,69906248 | 7,42286258 | 17,59234564 | 1 | Igual |  |  |  |
| 9,057252738 | 4,301033468 | 13,3126115 | 2 | Igual |  |  |  |
| 4,665290802 | 1,838175424 | 12,08141651 | 2 | Igual |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| **Iteración 3** |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Distancia 1 | Distancia 2 | Distancia 3 | Cluster final | Respecto al anterior |  |  |  |
| 9,081321939 | 3,578826623 | 7,856247514 | 2 | Igual |  |  |  |
| 0,745736179 | 6,997713912 | 17,4244261 | 1 | Igual |  |  |  |
| 6,840455886 | 3,264659247 | 10,19046736 | 2 | Igual |  |  |  |
| 11,92028797 | 7,347924877 | 5,988374153 | 3 | Igual |  |  |  |
| 4,790509325 | 1,994492417 | 12,79494529 | 2 | Igual |  |  |  |
| 20,99977162 | 14,30146846 | 6,971773447 | 3 | Igual |  |  |  |
| 19,58163446 | 13,85330286 | 3,151289419 | 3 | Igual |  |  |  |
| 16,67698524 | 11,05947558 | 1,706055392 | 3 | Igual |  |  |  |
| 5,755281514 | 8,527719508 | 19,36557835 | 1 | Igual |  |  |  |
| 13,73947835 | 15,53113003 | 18,17692562 | 1 | Igual |  |  |  |
| 2,217233061 | 7,277636979 | 18,02541608 | 1 | Igual |  |  |  |
| 2,017525257 | 7,210270453 | 17,92751028 | 1 | Igual |  |  |  |
| 1,003056553 | 6,225271078 | 16,74815288 | 1 | Igual |  |  |  |
| 2,247629136 | 6,811314117 | 17,59234564 | 1 | Igual |  |  |  |
| 9,882617186 | 5,414979224 | 13,3126115 | 2 | Igual |  |  |  |
| 5,800527773 | 1,180677771 | 12,08141651 | 2 | Igual |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| función objetivo 1 |  | función objetivo 2 |  | función objetivo 3 | |  |  |
| 27,72594005 |  | 15,43363528 |  | 17,81749241 |  |  |  |

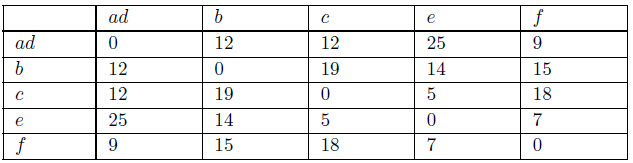
|  |  |  |
| --- | --- | --- |
| **Cluster** | | |
| 1 | 2 | 3 |
| 2,9,10,11,12,13,14 | 1,3,5,15,16 | 4,6,7,8 |

2. Para el ejemplo dado en la Sección 19.3.1, cuál sería la matriz de distancia después de cada una de las tres primeras fusiones si se usara el agrupamiento de enlaces completos en lugar de la agrupación de un solo enlace?

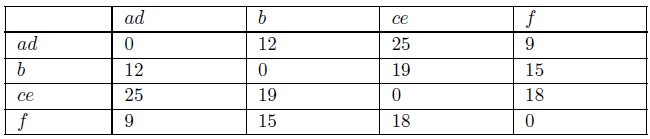
La matriz de distancias de esa sección es la siguiente:



Se junta los enlaces de la distancia más corta en este caso de a hacia d con una distancia de 3.



Ahora se agrupa la menor distancia de la tabla anterior y es 5 de c hacia e



Finalmente se selecciona la menor distancia asociada y es 9 de ad a f

