

Tarea 4

Wilmer Gonzalez

19 de junio de 2015

Contents

1	Presentación del problema	1
2	Descripción del set de datos	1
3	Respuestas	1
4		2

1 Presentación del problema

Responder todas las preguntas presentadas por *Abastos Crema* usando los métodos *hcluster* o *kmeans*

2 Descripción del set de datos

Muestras de laboratorio provistas por el cliente.

3 Respuestas

1. Grafica de los puntos contenidos en cada set de datos:
2. Sea una matriz de disimilaridades o distancias $D_{n \times n}$ es una matriz tal que su elemento i, j es una disimilaridad $d(ij)$ tal que $\forall i, j, k$:
 - $d(i, j) \geq 0$
 - $d(i, j) = 0$
 - $d(i, j) = d(j, i)$
 - $d(i, j) \leq d(i, k) + d(k, j)$

donde D es una matriz simetrica y su diagonal son 0.

Para la disimilaridad $d(i, j)$ representa una medida de la diferencia entre dos observaciones x_i y x_j en este caso usaremos la disimilaridad basada en distancia manhattan dado que ningun valor :

$$d(i, j) = \sqrt{\sum_{i=1}^p (x_{ic} - x_{cj})^2}$$

$$d(i, j) = \sum_{i=1}^p \|(x_{ic} - x_{cj})\|$$

especificamente el criterio de vecino mas cercano expresado como :

$$d_{UV} = \min(d_{ij}) : i \in U, j \in V$$

ya que, los conglomerados formados por este data set no poseen formas estrictamente esfericas y por lo tanto se ajustarian mas (teoricamente) las comparaciones individuales de vecino mas cernano.

3. Para cada dataset se generaron los siguientes dendogramas:

4