

Inferencia Estadística y Reconocimiento de Patrones

UNaB, Lic. Cs. de Datos, 2021 cuat. 2

Sebastián Pedersen (sebastian.pedersen (at) unab (punto) edu (punto) ar)

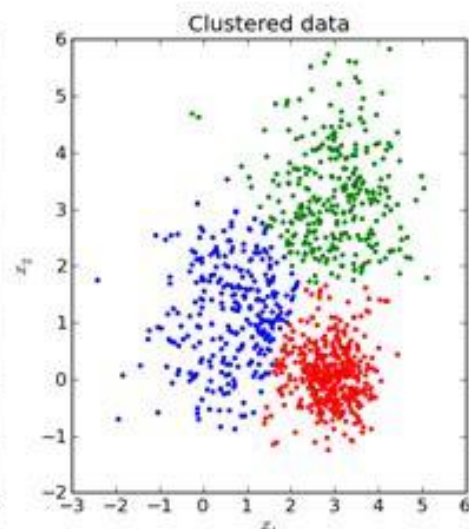
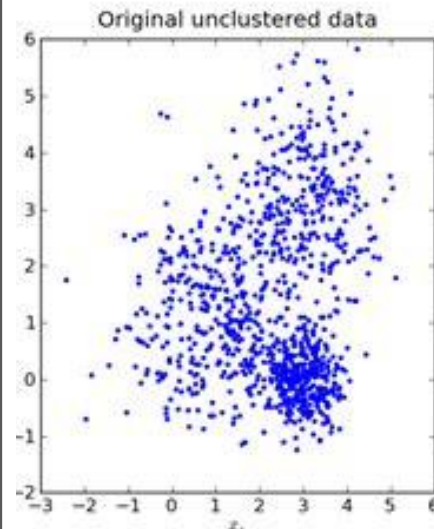
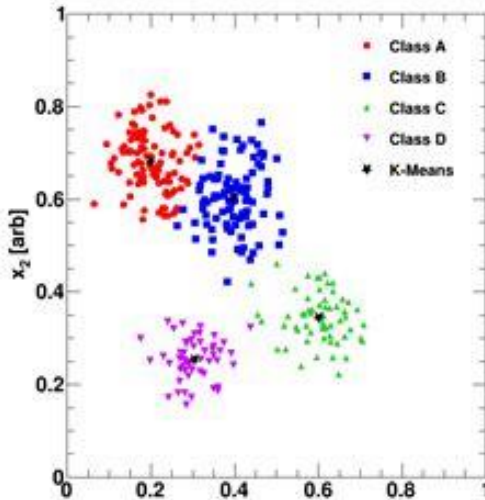
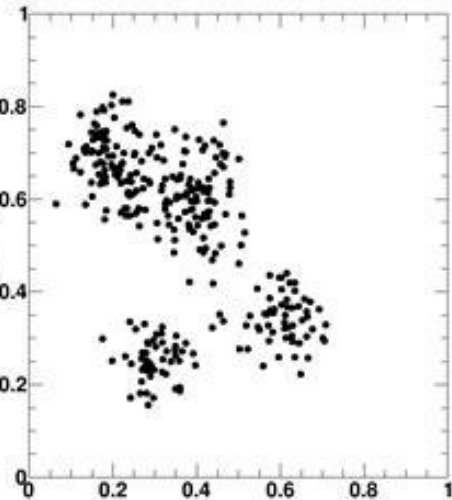
Clustering jerárquico aglomerativo

(Clasificación NO supervisada)

Clasificación No Supervisada

Se tiene un conjunto de datos SIN clasificar.

Objetivo: extraer agrupamientos o estructuras internas, en lo posible con interpretación.



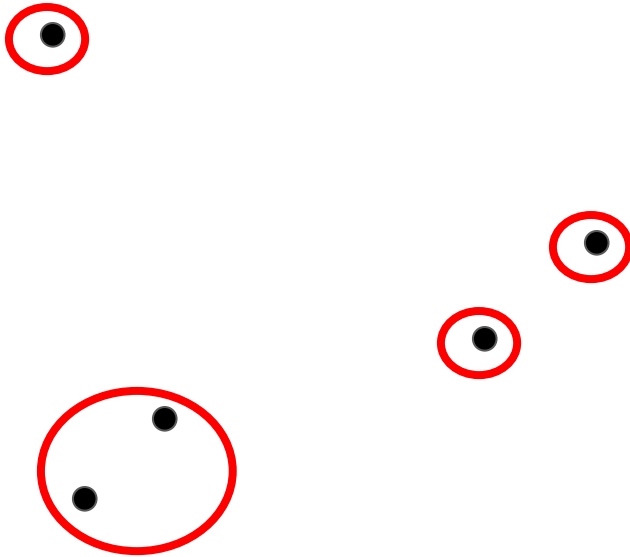
Jerárquico aglomerativo paso 1: todos los datos son un cluster



En este caso hay 5 clusters.



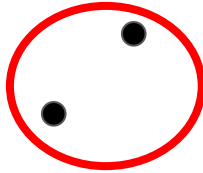
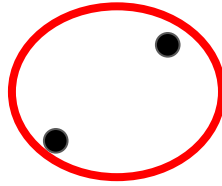
Jerárquico aglomerativo paso 2: agrupar los dos clusters más cercanos entre sí.



Ahora hay 4 clusters.

Si la cantidad de clusters deseada es 4, el algoritmo para acá.

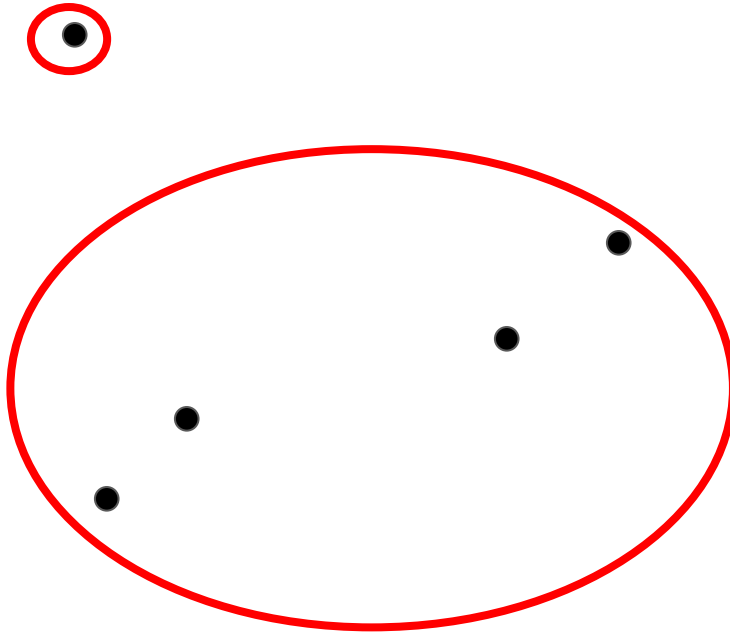
Jerárquico aglomerativo paso 2 (de nuevo): agrupar los dos clusters más cercanos entre sí.



Ahora hay 3 clusters.

Si la cantidad de clusters deseada es 3, el algoritmo para acá.

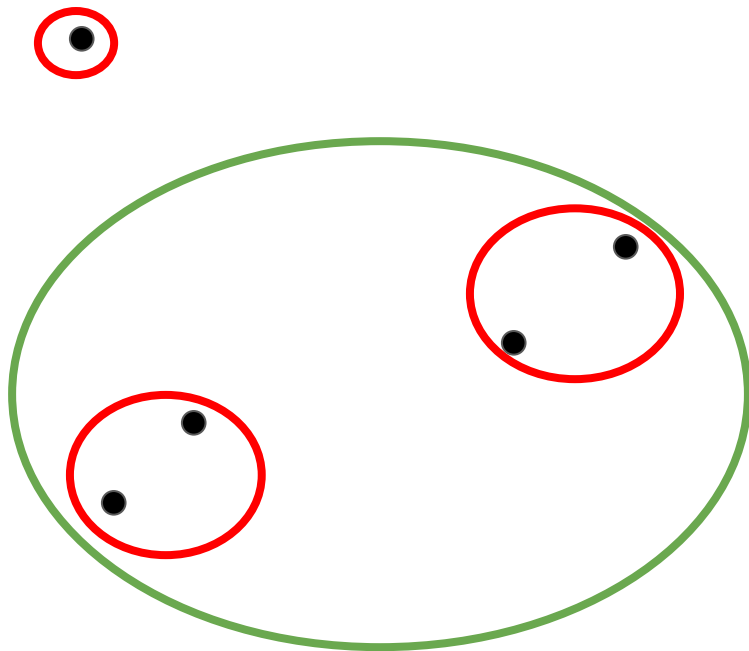
Jerárquico aglomerativo paso 2 (de nuevo): agrupar los dos clusters más cercanos entre sí.



Ahora hay 2 clusters.

El algoritmo para, pues
menos de 2 no tiene
sentido.

Jerárquico aglomerativo paso 2: detalle importante.



Observar que acá unió dos clusters de muchos puntos (el verde unió dos rojos).

¿Cómo medir la distancia entre clusters, o entre clusters y un punto?

Varias formas: ward, single, average, complete.

(Queda en el tintero para la vez que viene).

Referencias

- Hastie, Tibshirani, Introduction to Statistical Learning, sección 12.4.2
- Chan, Análisis Inteligente de Datos, sección 10.2.3