

Inferencia Estadística y Reconocimiento de Patrones

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

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

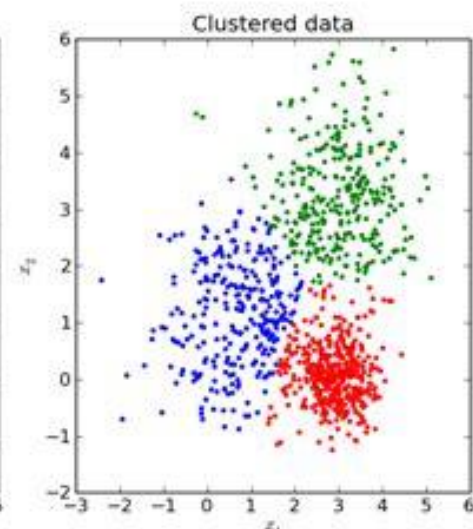
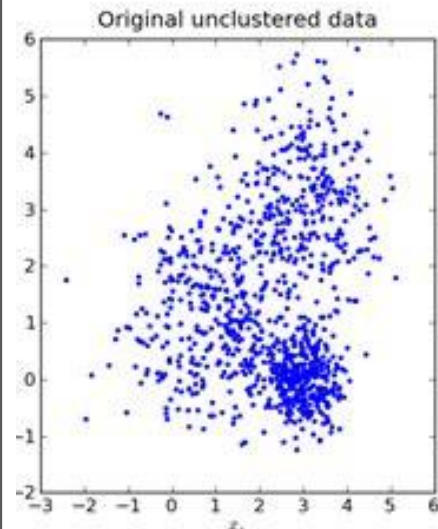
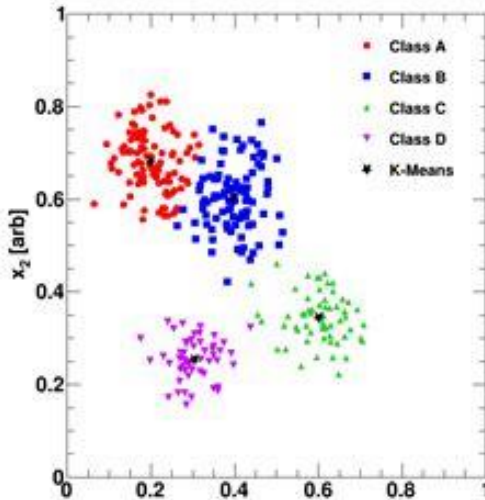
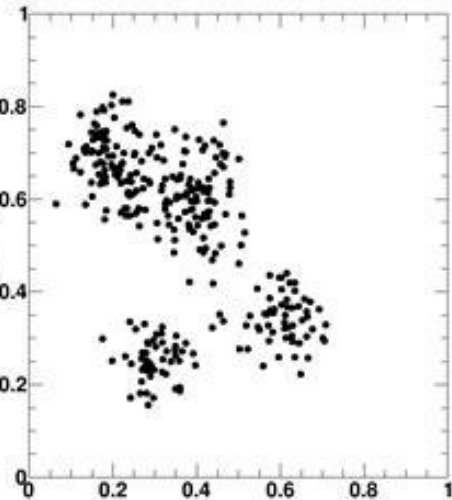
Clustering por k-means

(Clasificación NO supervisada)

Clasificación No Supervisada

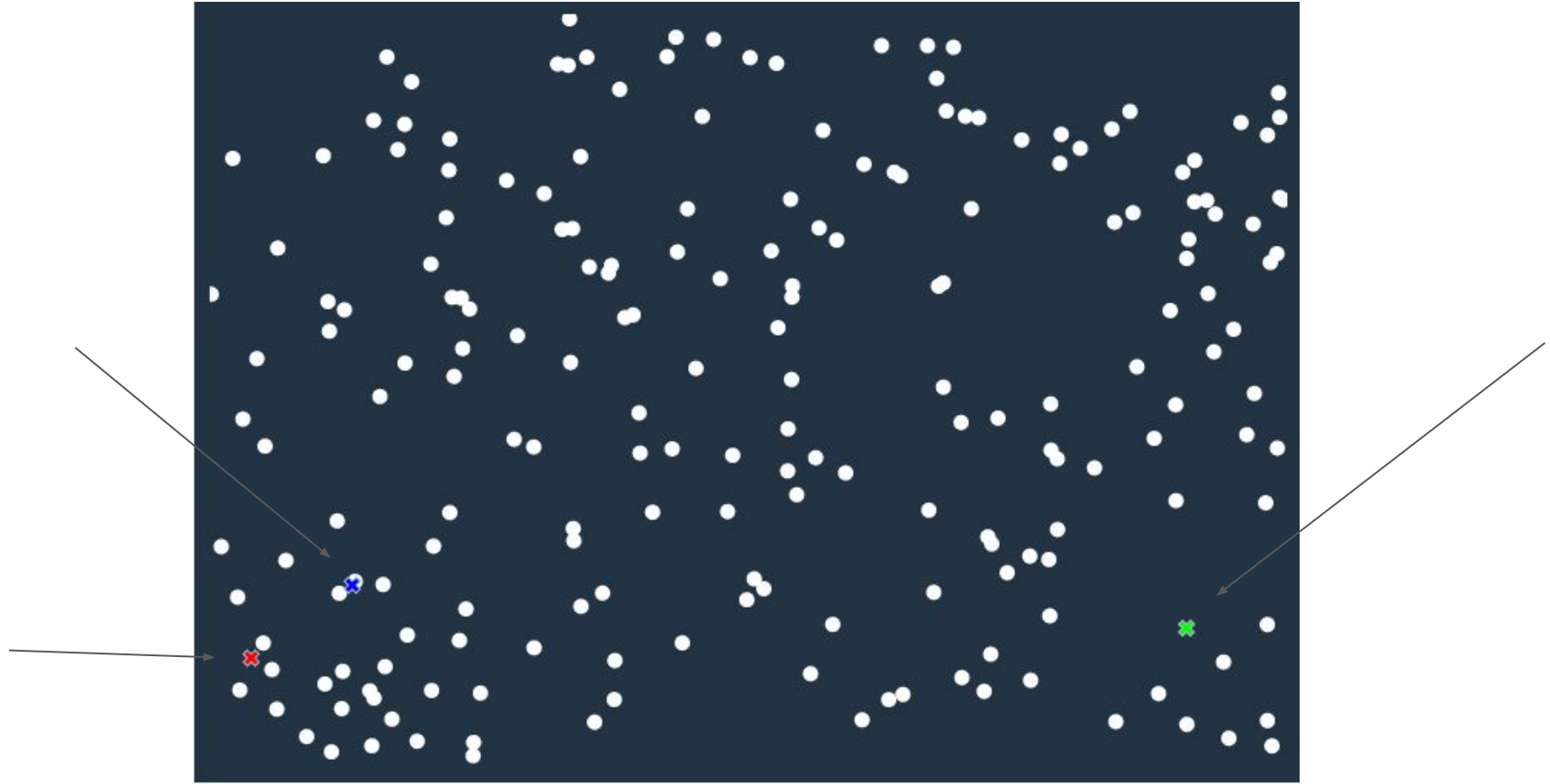
Se tiene un conjunto de datos SIN clasificar.

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



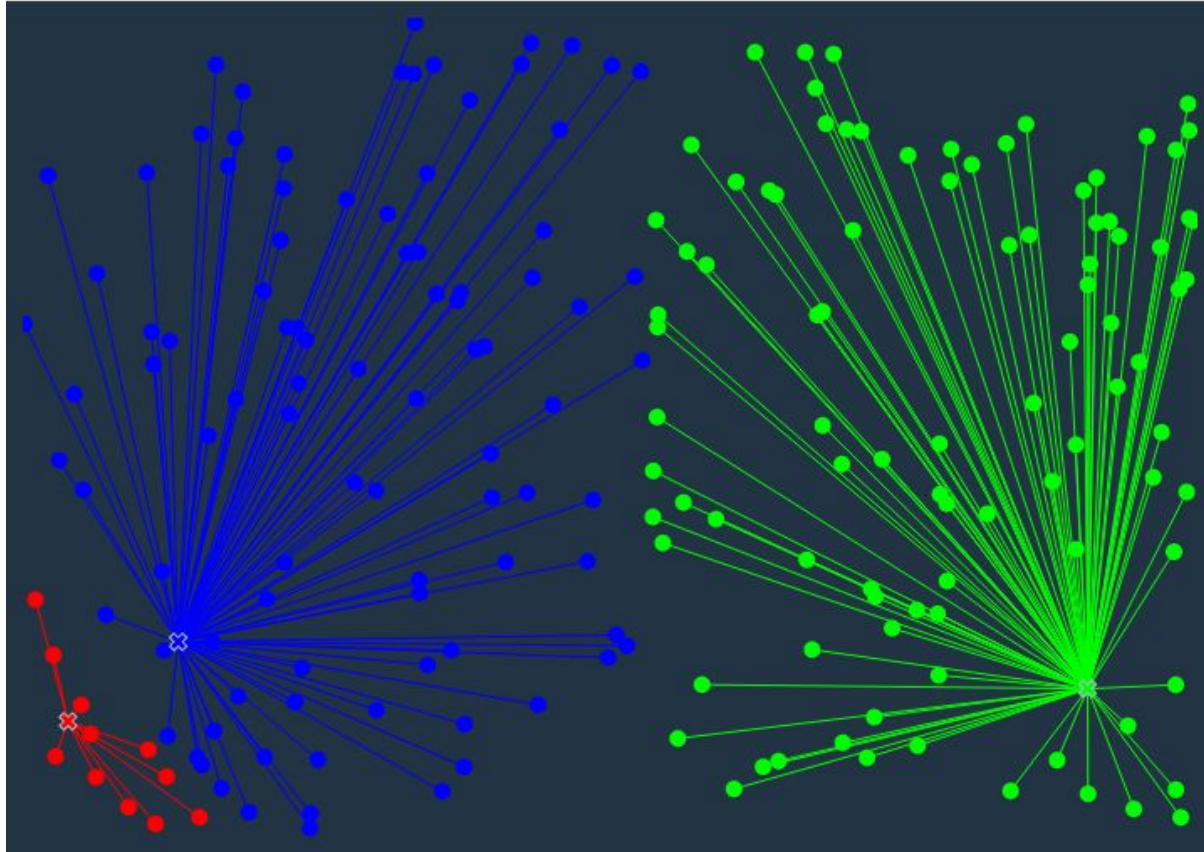
k-means paso 1: inicialización

Se tiene un conjunto de datos SIN clasificar. Se elijen al azar k puntos.



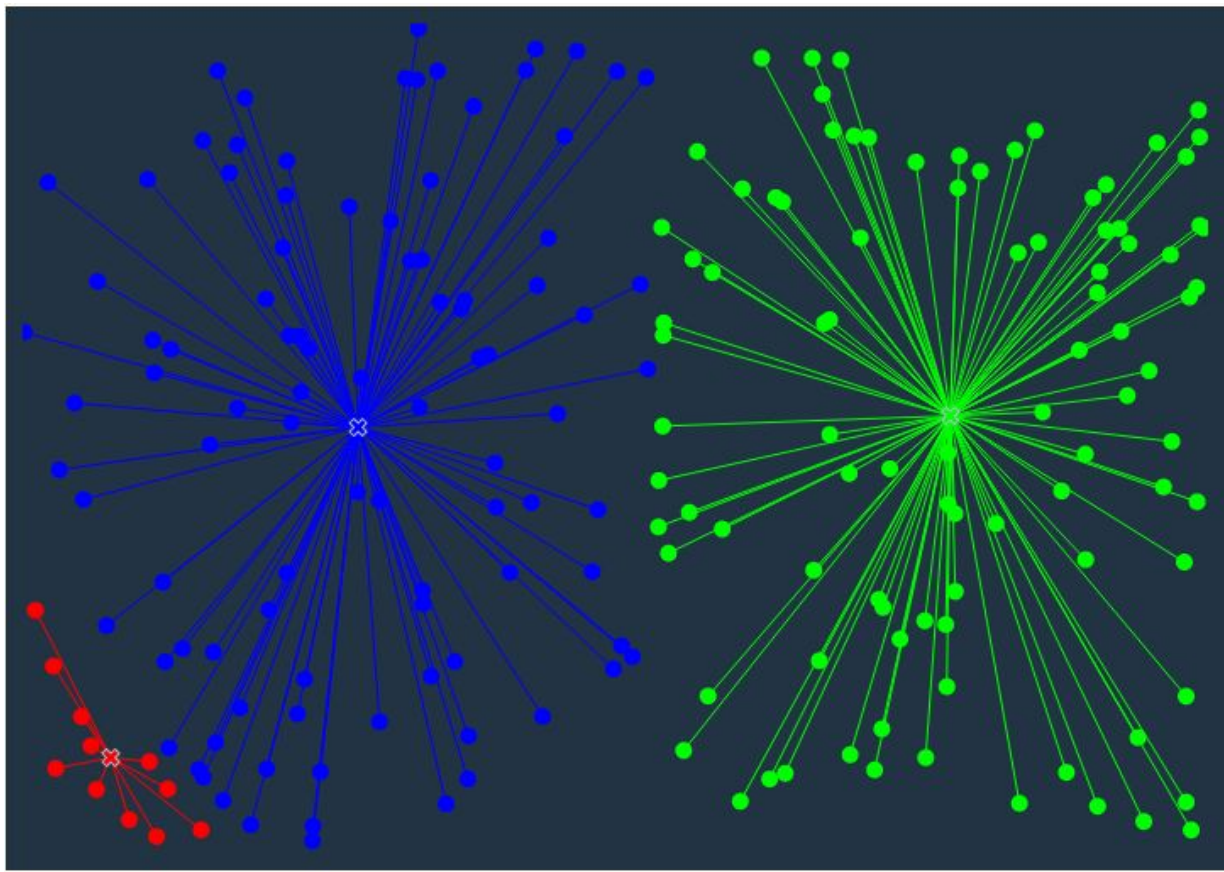
k-means paso 2: asignación

Asignar cada dato (punto) al cluster más cercano (cruz).



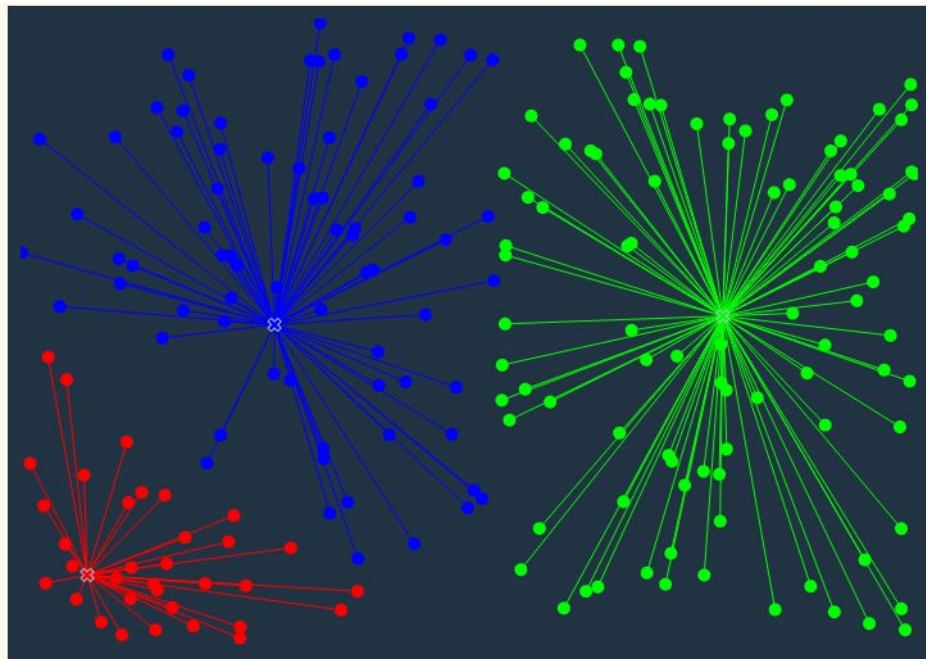
k-means paso 3: recalcular centroides.

Recalcular cada centro (cruz) como la media de sus datos.

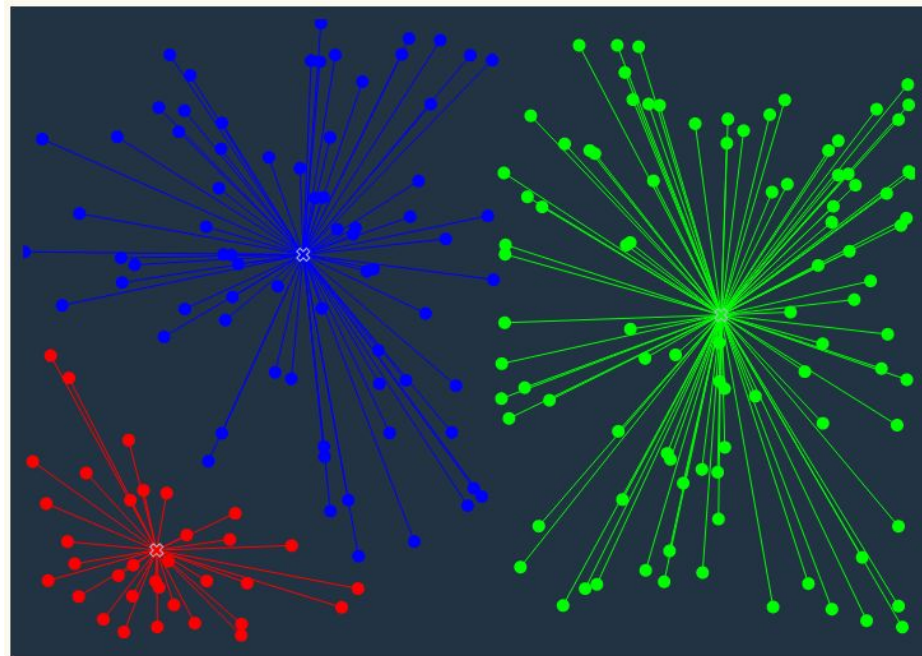


k-means paso 4: repetir 2 y 3 hasta que no haya cambios.

PASO 2: asignación (de datos a clusters)



PASO 3: recalcular centroides.



Referencias

- Hastie, Tibshirani, Introduction to Statistical Learning, sección 12.4.1
- Bishop, Pattern Recognition and Machine Learning, sección 9.1
- Chan, Análisis Inteligente de Datos, sección 10.2.7
- Visualización interactiva k-means: <https://kkevsterrr.github.io/K-Means/>