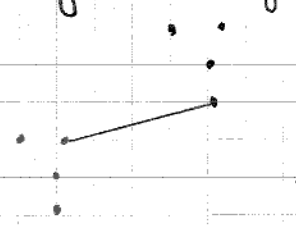
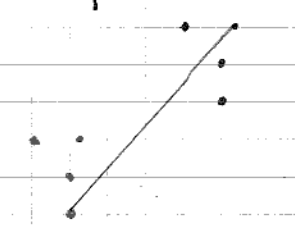
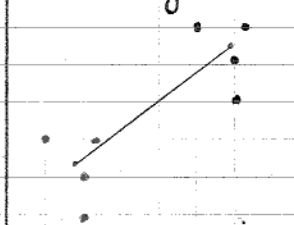
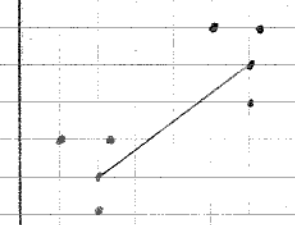


Clustering.

Clustering Jerárquico

- 1) Partimos de que cada punto es un cluster.
- 2) En cada paso unimos los dos clusters más cercanos.
- 3) Repetir hasta que haya un único cluster, u obtenemos la cantidad que queríamos.

Distancias entre clusters

Single-linkage	Complete-linkage	Average-link	Median-link
			
Minima conexión entre clusters	Máxima conexión entre clusters	"Centro de masa"	Conexión mediana entre clusters.

Optimización con LSH para Single-linkage

- Es una aproximación.
- En cada bucket habrá como máximo un punto de cada cluster.
- Los clusters en un mismo bucket son candidatos.

K-Means

- Dado un set de puntos en \mathbb{R}^d
- Particionar los puntos en K sets ($K \leq d$) de forma tal de minimizar la sumatoria de las distancias de cada punto al centro del cluster.

$$\sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|.$$

NOTA: Es un problema ~~NP-hard~~ NP-hard de resolver óptimamente pero existen formas de aproximarlos.

Algoritmo:

- 1) Elegir k puntos, uno para cada cluster.
a. Inicialmente, los ~~no~~ centroides de cada cluster son esos puntos.
- 2) Asignarle a cada punto el cluster con centenoide más cercano.
- 3) Recalcular los centroides como el promedio de los puntos.
- 4) Volver a 2.

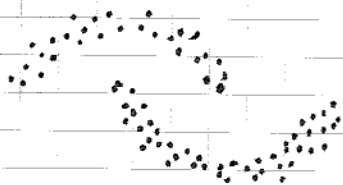
El problema con este algoritmo es que es muy dependiente de los puntos iniciales (si los que tomaron son muy malos entonces el algoritmo ~~no~~ tarda mucho).

K-Means ++

- 1) Elegir un punto al azar como centro de
- 2) Calcular la distancia de cada punto con los centros y quedarse con la mínima.
- 3) Calcular la probabilidad de cada punto como la distancia dividida por la suma de todas las distancias de los puntos.
- 4) Elegir un nuevo punto al azar, ~~utilizando~~ utilizando la probabilidad calculada.
- 5) Repetir hasta tener los centros.

Clustering espectral

Busca mejorar K-means para que pueda encontrar clusters no lineales como:



Funciona primero aplicando la Transformación de Laplacian eigenmap y después aplicando K-means.

Editor DBScan y TDBScan.