

Aprendizaje no supervisado

Algoritmo de K-medias

Julio Weissman Vilanova

Licenciatura en Ciencias de la Computación
Universidad de Sonora

Curso Inteligencia Artificial

Plan del curso

Aprendizaje no supervisado

El algoritmo de K-Medias

Propiedades de K-Medias

Resumen

¿Aprendizaje?

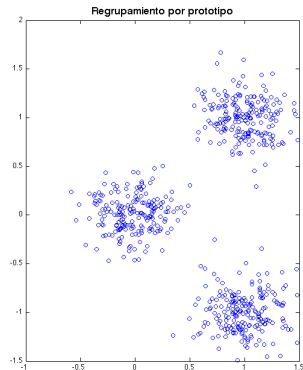
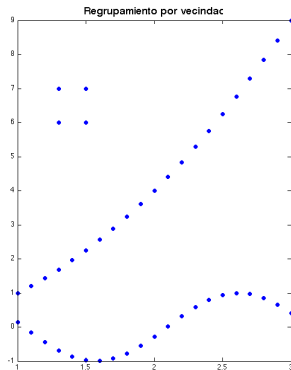
- ▶ El aprendizaje **supervisado** utiliza un conjunto de datos de aprendizaje previamente clasificado. Aprendizaje con maestro.
- ▶ El aprendizaje **no supervisado** utiliza un conjunto de datos sin clasificar. Descubrimiento de conocimiento en bases de datos (KDD).
- ▶ El aprendizaje **reforzado** utiliza la interacción con el medio para establecer una política de comportamiento. Aprendizaje con crítico.

Aprendizaje no supervisado

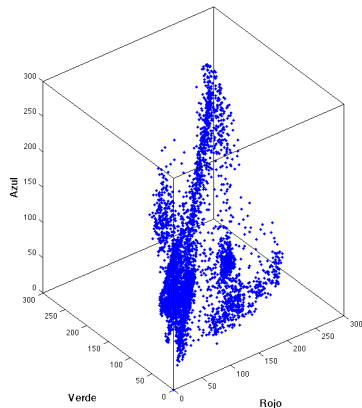
- ▶ Objetos descritos por m atributos: $x_i = [x_{i,1}, \dots, x_{i,m}]$.
- ▶ Conjunto de aprendizaje: $X = \{x_1, \dots, x_n\}$.
- ▶ Determinar $y_i = f(x_i, \theta)$ tal que $y_i \in \{P_1, \dots, P_K\}$, tal que todos los objetos para los cuales $f(x, \theta) = P_j$ sean similares en algún sentido.
- ▶ ¡La **interpretabilidad** de los grupos (*clusters*) es esencial!

Formas de hacer agrupamiento

- ▶ Jerárquico.
- ▶ Vecindad.
- ▶ Prototipo ideal.

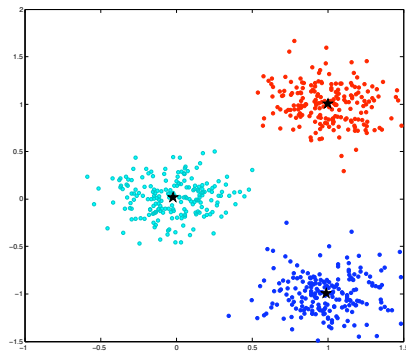


No siempre es sencillo encontrar los prototipos



Idea básica

Un objeto pertenece a un grupo si se encuentra más próximo al prototipo de esa clase que al de cualquier otra.



Encontrar los prototipo ideales tales que los datos estén lo más compactos entre sí en cada grupo.

Idea básica

- ▶ Prototipo de clase $c_j = [c_{j,1}, \dots, c_{j,m}]$.
- ▶ Los parámetros de aprendizaje son $\theta = c_1, \dots, c_K$.
- ▶ Distancia euclidiana entre un objeto y un prototipo:

$$d(x_i, c_j) = \sqrt{\sum_{l=1}^m (x_{i,l} - c_{j,l})^2}.$$

- ▶ Minimizar:

$$J(\theta) = \sum_{i=1}^n \min_k (d(x_i, c_k)).$$

- ▶ ¡Problema NP completo!

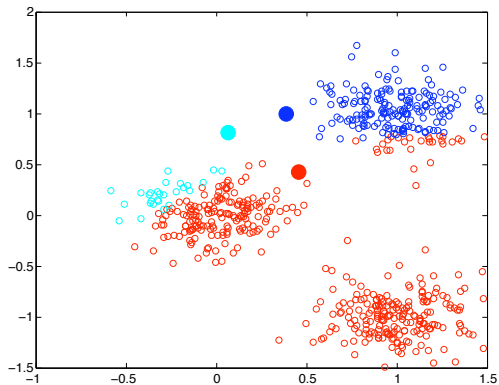
El algoritmo de las K-medias

Entrada: $\{x_1, \dots, x_n\}$, K , m .

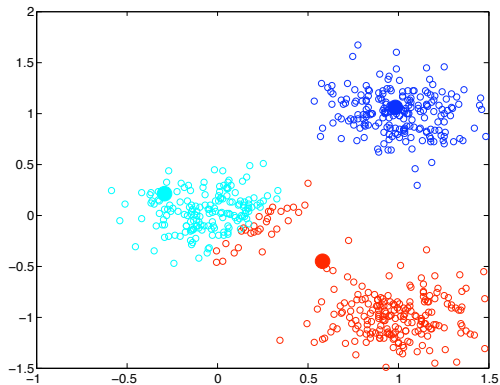
Salida: $\{c_1, \dots, c_K\}$.

- 1: Inicializa $\{c_1, \dots, c_K\}$ con valores aleatorios.
- 2: **repetir**
- 3: **para** i de 1 a n **hacer**
- 4: $x_i \in P_k$ si $\min_k(d(x_i, c_k))$.
- 5: **fin para**
- 6: **para** k de 1 a K **hacer**
- 7: **para** j de 1 a m **hacer**
- 8:
$$c_{k,j} = \frac{\sum_{x_i \in P_k} x_{i,j}}{||P_k||}$$
- 9: **fin para**
- 10: **fin para**
- 11: **hasta** no varíe $c_{k,j}, \forall k, j$

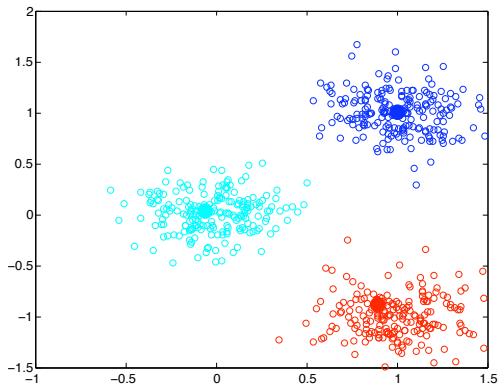
Un ejemplo simple



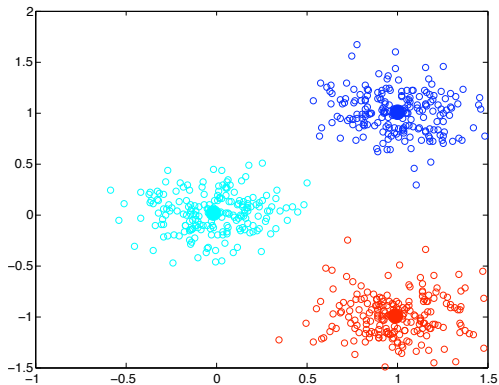
Un ejemplo simple



Un ejemplo simple



Un ejemplo simple



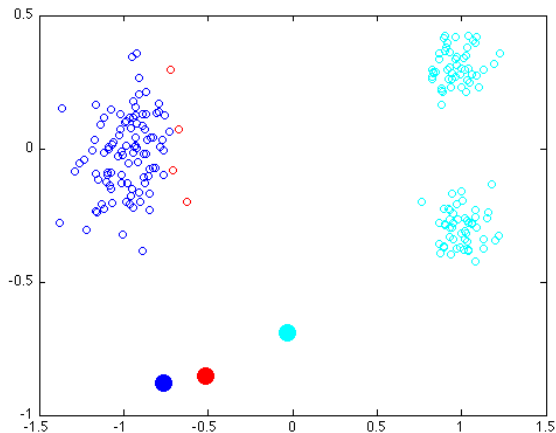
Aplicación a segmentación de imágenes



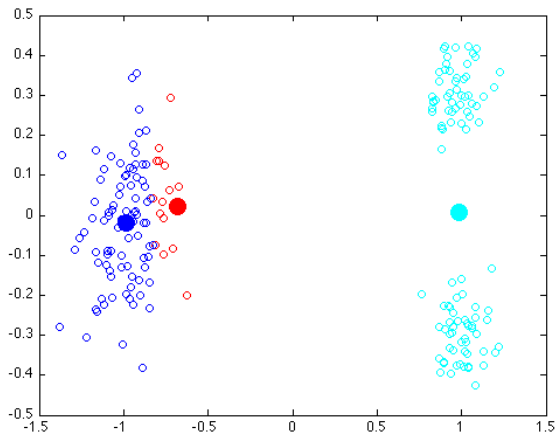
Propiedades del algoritmo

- ▶ Basado en gradiente descendiente.
- ▶ Convergencia asegurada.
- ▶ En general obtiene mínimos locales.
- ▶ ¡Cuidado con la inicialización!

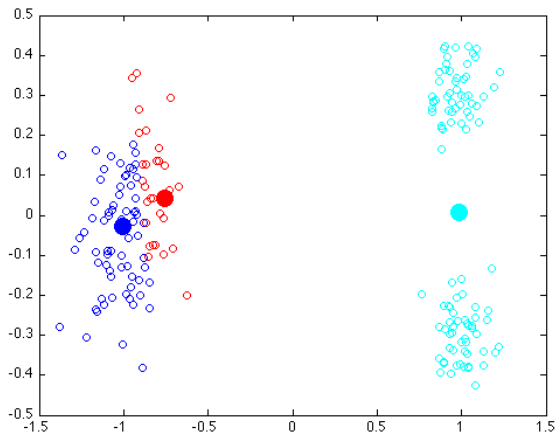
Un ejemplo simple



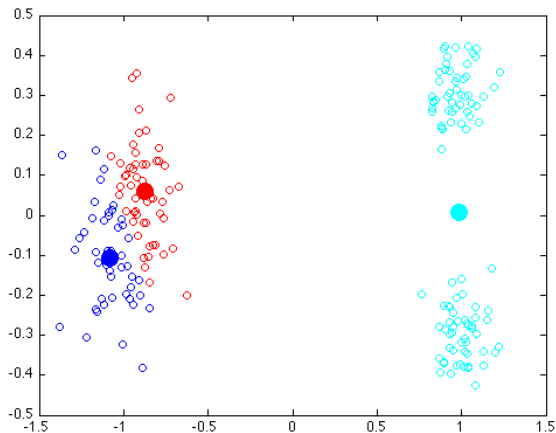
Un ejemplo simple



Un ejemplo simple



Un ejemplo simple



Propiedades del algoritmo

Ventajas

- ▶ Sencillez
- ▶ Algoritmo eficiente.
- ▶ Versión por lotes y versión en línea.

Desventajas

- ▶ Depende fuertemente de la inicialización.
- ▶ Número de clases *a priori*.
- ▶ Sólo particiones del espacio por hiperplanos.

Extensiones de K-Medias

- ▶ Métodos de inicialización de centros.
- ▶ Selección de número de clases.
- ▶ Valores perdidos.
- ▶ Interpretación de los prototipos.
- ▶ Uso de distancias adaptables (Mahalanobis).
- ▶ Pertenencia difusa a los grupos (*Fuzzy C-Means*).

Resumen

Hoy vimos:

- ▶ Qué es el aprendizaje no supervisado.
- ▶ El algoritmo de K-Medias y su aplicación.
- ▶ Precauciones con el algoritmo.

Ahora pueden:

- ▶ Programar el algoritmo K-Medias.
- ▶ Aplicarlo a tablas de datos y segmentación de imágenes.

Para más información



R Duda, P. Hart y D. Stork.

Pattern Classification.

Wiley Interscience, 2001 (2ª Impresión).



D. McKay.

Information Theory, Inference and Learning Algorithms.

Cambridge University Press.

<http://www.inference.phy.cam.ac.uk/mackay/itprnn/ps/284.292.pdf>