Clustering

Mezclas de Gaussianas

Christian Oliva Moya

Dpto. de Ingeniería Informática, Escuela Politécnica Superior

Universidad Autónoma de Madrid

28049 Madrid, Spain

v1 - 21/02/2024

Índice

1. Introducción
2. Teorema de Bayes
3. Teorema de Bayes para E-M
4. Algoritmo E-M
    a. Paso 1: Expectation
    b. Paso 2: Maximization
    c. Notebook

Introducción (1)

- Expectation-Maximization (E-M) describe probabilísticamente los grupos

  K-means asume que los datos se organizan en K grupos según los centroides (medias).

  E-M asume K distribuciones estadísticas (Gaussianas: medias + varianzas).

- Es un algoritmo estocástico (iterativo), como K-means.

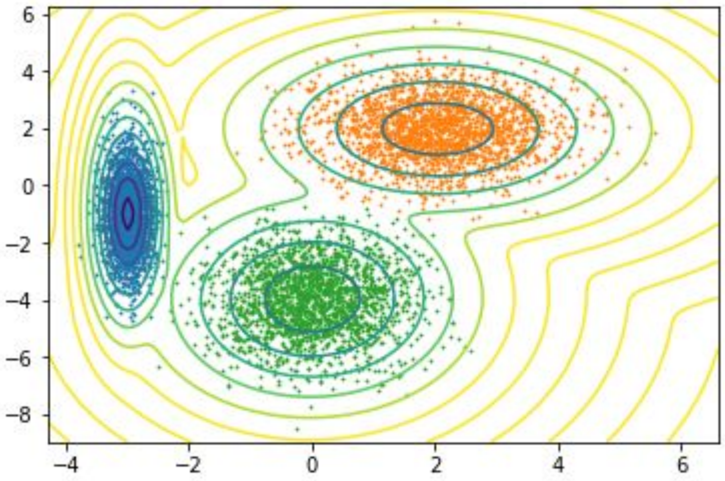
  Comienza con parámetros aleatorios e itera hasta converger.

- Es un algoritmo difuso (soft).

  Un punto puede pertenecer a varios clusters (probabilidades).

Introducción (2)

- Ejemplo de ejecución:
    - Se forman K=3 Gaussianas (solapadas)
    - Cada una con media y varianza diferentes



Introducción (3)

- Nomenclatura:
    - Conjunto de n elementos $X = (x_1, x_2, \ldots, x_n)$ con dimensión 1
    - Asumimos que existen k Gaussianas $G = (G_1, G_2, \ldots, G_k)$

      Con sus respectivos parámetros $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ y $\sigma^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$

Teorema de Bayes

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

- Donde:
  - $P(G_j \mid x_i)$ es la probabilidad de que x_i pertenezca a G_j → Probabilidad a posteriori
  - $P(G_j)$ es la probabilidad de pertenecer a la gaussiana G_j → Probabilidad a priori
  - $P(x_i)$ es la probabilidad de seleccionar el elemento x_i → Evidencia
  - $P(x_i \mid G_j)$ es la verosimilitud (likelihood), que sigue una distribución gaussiana:

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2})$$

Teorema de Bayes para E-M (1)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2})$$

- Asumimos los siguientes puntos:
  - $P(G_j)$ es la probabilidad de pertenecer a la gaussiana G_j → Probabilidad a priori

    No conocemos los prioris → $P(G_j) = 1 / K$ para todas las gaussianas.
  - $P(x_i)$ es la probabilidad de seleccionar el elemento x_i → Evidencia

    La probabilidad de cada punto es igual para todos → $P(x_i) = 1 / n$ para todos los puntos.

Teorema de Bayes para E-M (2)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2})$$

- Ejemplo:

  $X = (x_1 = 3, \; x_2 = 2, x_3 = -3, x_4 = -2)$ y $K = 2$

  $G_1 = (\mu_1 = 2.0, \sigma_1 = 1.0)$ y $G_2 = (\mu_2 = -2.0, \sigma_2 = 1.0)$

  Calculamos los prioris y las evidencias:

Teorema de Bayes para E-M (2)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2})$$

- Ejemplo:

  $X = (x_1 = 3,\ x_2 = 2, x_3 = -3, x_4 = -2)$ y $K = 2$

  $G_1 = (\mu_1 = 2.0, \sigma_1 = 1.0)$ y $G_2 = (\mu_2 = -2.0, \sigma_2 = 1.0)$

  Calculamos los prioris y las evidencias:

  $P(G_1) = P(G_2) = 1 / 2 = 0.5$

  $P(x_1) = P(x_2) = P(x_3) = P(x_4) = 1 / 4 = 0.25$

  Calculamos verosimilitudes G1: Calculamos verosimilitudes G2:

Teorema de Bayes para E-M (2)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2})$$

- Ejemplo:

  $X = (x_1 = 3, \; x_2 = 2, x_3 = -3, x_4 = -2)$ y $K = 2$

  $G_1 = (\mu_1 = 2.0, \sigma_1 = 1.0)$ y $G_2 = (\mu_2 = -2.0, \sigma_2 = 1.0)$

  Calculamos los prioris y las evidencias:

  $P(G_1) = P(G_2) = 1 / 2 = 0.5$

  $P(x_1) = P(x_2) = P(x_3) = P(x_4) = 1 / 4 = 0.25$

  Calculamos verosimilitudes G1:

  $P(x_1 \mid G_1) = \ldots = 0.241971$

  $P(x_2 \mid G_1) = \ldots = 0.398942$

  $P(x_3 \mid G_1) = \ldots = 0.000001$

  $P(x_4 \mid G_1) = \ldots = 0.000134$

  Calculamos verosimilitudes G2:

  $P(x_1 \mid G_2) = \ldots = 0.000001$

  $P(x_2 \mid G_2) = \ldots = 0.000134$

  $P(x_3 \mid G_2) = \ldots = 0.241971$

  $P(x_4 \mid G_2) = \ldots = 0.398942$

Calculamos las probabilidades:

Teorema de Bayes para E-M (2)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}}exp(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2})$$

- Ejemplo:

  $X = (x_1 = 3,\ x_2 = 2, x_3 = -3, x_4 = -2)$ y $K = 2$

  $G_1 = (\mu_1= 2.0, \sigma_1=1.0)$ y $G_2 = (\mu_2= -2.0, \sigma_2=1.0)$

  Calculamos los prioris y las evidencias:

  $P(G_1) = P(G_2) = 1 / 2 = 0.5$

  $P(x_1) = P(x_2) = P(x_3) = P(x_4) = 1 / 4 = 0.25$

  Calculamos verosimilitudes G1:

  $P(x_1 | G_1) = \ldots = 0.241971$

  $P(x_2 | G_1) = \ldots = 0.398942$

  $P(x_3 | G_1) = \ldots = 0.000001$

  $P(x_4 | G_1) = \ldots = 0.000134$

  Calculamos verosimilitudes G2:

  $P(x_1 | G_2) = \ldots = 0.000001$

  $P(x_2 | G_2) = \ldots = 0.000134$

  $P(x_3 | G_2) = \ldots = 0.241971$

  $P(x_4 | G_2) = \ldots = 0.398942$

  Calculamos las probabilidades:

  $P(G_1 | x_1) = \ldots = 0.483941$

  $P(G_1 | x_2) = \ldots = 0.797884$

  $P(G_1 | x_3) = \ldots = 0.000003$

  $P(G_1 | x_4) = \ldots = 0.000268$

  $P(G_2 | x_1) = \ldots = 0.000003$

  $P(G_2 | x_2) = \ldots = 0.000268$

  $P(G_2 | x_3) = \ldots = 0.483941$

  $P(G_2 | x_4) = \ldots = 0.797884$

Teorema de Bayes para E-M (3)

- Ejemplo:
  - Asignamos al cluster con mayor probabilidad:

$P(G_1 \mid x_1) = \ldots = 0.483941 \quad \rightarrow \quad 0.999999$

$P(G_2 \mid x_1) = \ldots = 0.000003 \quad \rightarrow \quad 0.000001$

$P(G_1 \mid x_2) = \ldots = 0.797884 \quad \rightarrow \quad 0.999664$

$P(G_2 \mid x_2) = \ldots = 0.000268 \quad \rightarrow \quad 0.000336$

$P(G_1 \mid x_3) = \ldots = 0.000003 \quad \rightarrow \quad 0.000001$

$P(G_2 \mid x_3) = \ldots = 0.483941 \quad \rightarrow \quad 0.999999$

$P(G_1 \mid x_4) = \ldots = 0.000268 \quad \rightarrow \quad 0.000336$

$P(G_2 \mid x_4) = \ldots = 0.797884 \quad \rightarrow \quad 0.999664$

Teorema de Bayes para E-M (4)

$$P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

- Hemos asumido que:
  - No conocemos los prioris → $P(G_j) = 1 / K$ para todas las gaussianas.
  - La probabilidad de cada punto es igual para todos → $P(x_i) = 1 / n$ para todos los puntos.

- ¿Podemos inferir lo siguiente?

$$P(G_j|x_i) = P(x_i|G_j)$$

OJO!! Solamente en E-M para clustering!!

Teorema de Bayes para E-M (5)

- Ejemplo:

$$
\begin{array}{lcccl}
P(G_1 \mid x_1) = \ldots = 0.483941 & \rightarrow & 0.999999 & \leftarrow & P(x_1 \mid G_1) = \ldots = 0.241971 \\
P(G_2 \mid x_1) = \ldots = 0.000003 & \rightarrow & 0.000001 & \leftarrow & P(x_1 \mid G_2) = \ldots = 0.000001 \\
P(G_1 \mid x_2) = \ldots = 0.797884 & \rightarrow & 0.999664 & \leftarrow & P(x_2 \mid G_1) = \ldots = 0.398942 \\
P(G_2 \mid x_2) = \ldots = 0.000268 & \rightarrow & 0.000336 & \leftarrow & P(x_2 \mid G_2) = \ldots = 0.000134 \\
P(G_1 \mid x_3) = \ldots = 0.000003 & \rightarrow & 0.000001 & \leftarrow & P(x_3 \mid G_1) = \ldots = 0.000001 \\
P(G_2 \mid x_3) = \ldots = 0.483941 & \rightarrow & 0.999999 & \leftarrow & P(x_3 \mid G_2) = \ldots = 0.241971 \\
P(G_1 \mid x_4) = \ldots = 0.000268 & \rightarrow & 0.000336 & \leftarrow & P(x_4 \mid G_1) = \ldots = 0.000134 \\
P(G_2 \mid x_4) = \ldots = 0.797884 & \rightarrow & 0.999664 & \leftarrow & P(x_4 \mid G_2) = \ldots = 0.398942
\end{array}
$$

Algoritmo E-M (1)

- Inicializar de forma aleatoria las gaussianas. Por ejemplo:

  Los K puntos de partida son puntos del dataset seleccionados al azar.

  Se calculan las K medias y varianzas de los datos más cercanos a cada punto de partida.

- Mientras no converja el algoritmo:
  - Paso 1: Expectation (asignación a grupos)
  - Paso 2: Maximization (actualización de medias y varianzas)

Algoritmo E-M (2) – Paso 1

- Expectation: Asignar cada elemento x_i al grupo S_j cuya verosimilitud sea máxima.

$$P(x_i|G_i) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2})$$

$$\theta_{ij} = P(G_j|x_i) = \frac{P(x_i|G_j)P(G_j)}{P(x_i)}$$

$$S_j = S_j \cup \{ \, argMax(\, \theta_{ij}) \text{ for each } j \text{ in range}(K) \, \}$$

Algoritmo E-M (2) – Paso 2

- Maximization. Recalcular los parámetros para cada G_j

$$\mu_j^{(t+1)} = \frac{\sum \theta_{ij}^{(t)} x_i}{\sum \theta_{ij}^{(t)}} \qquad \sigma_j^{2:(t+1)} = \frac{\sum \theta_{ij}^{(t)} (x_i - \mu_j^{(t+1)})^2}{\sum \theta_{ij}^{(t)}}$$

Algoritmo E-M (3) – Notebook

- Algoritmo E-M en python:
  - Notebook 04_ejemplo_EM.ipynb