

El Algoritmo EM y el modelo de mezcla de Normales.

Dimayuga Neyva
Moreno Alejandra
Sánchez Edoardo

UAM-I

Una travesía por la Ciencia de Datos para matemáticos y no matemáticos.

MCMAI

Un poco de nosotros:

- **Dimayuga Neyva:**

Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.

- **Moreno Alejandra:**

Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.

- **Sánchez Edoardo:**

Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

Un poco de nosotros:

- **Dimayuga Neyva:**
Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.
- **Moreno Alejandra:**
Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.
- **Sánchez Edoardo:**
Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

Un poco de nosotros:

- **Dimayuga Neyva:**
Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.
- **Moreno Alejandra:**
Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.
- **Sánchez Edoardo:**
Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

Un poco de nosotros:

- **Dimayuga Neyva:**
Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.
- **Moreno Alejandra:**
Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.
- **Sánchez Edoardo:**
Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

Un poco de nosotros:

- **Dimayuga Neyva:**
Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.
- **Moreno Alejandra:**
Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.
- **Sánchez Edoardo:**
Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

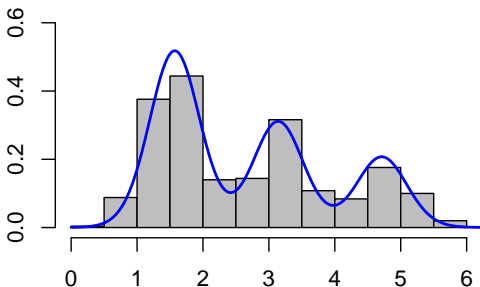
Un poco de nosotros:

- **Dimayuga Neyva:**
Inferencia Variacional para Datos Circulares: Un Enfoque Bayesiano.
- **Moreno Alejandra:**
Análisis Bayesiano de un Modelo No-paramétrico Basado en Distribuciones von Mises.
- **Sánchez Edoardo:**
Técnicas de muestreo paralelo en distribuciones finales Bayesianas: El caso de distribuciones para datos circulares.

¿Qué es el modelo de mezclas?

$$f(x) = \sum_{j=1}^{\infty} \omega_j k_j(x|\underline{\theta}_j)$$

Model 1, m = 500



Modelo de mezclas infinitas

Un pequeño paso para tratar de entender el modelo de mezclas infinito:

- Cluster (Agrupamiento de datos)
- Modelo de Mezclas finito

$$f(x) = \sum_{j=1}^M \omega_j k_j(x|\underline{\theta}_j), \quad M \in \mathbb{N}.$$

Modelo de mezclas infinitas

Un pequeño paso para tratar de entender el modelo de mezclas infinito:

- Cluster (Agrupamiento de datos)
- Modelo de Mezclas finito

$$f(x) = \sum_{j=1}^M \omega_j k_j(x|\underline{\theta}_j), \quad M \in \mathbb{N}.$$

Modelo de mezclas infinitas

Un pequeño paso para tratar de entender el modelo de mezclas infinito:

- Cluster (Agrupamiento de datos)
- Modelo de Mezclas finito

$$f(x) = \sum_{j=1}^M \omega_j k_j(x|\underline{\theta}_j), \quad M \in \mathbb{N}.$$

Modelo de mezclas infinitas

Un pequeño paso para tratar de entender el modelo de mezclas infinito:

- Cluster (Agrupamiento de datos)
- Modelo de Mezclas finito

$$f(x) = \sum_{j=1}^M \omega_j k_j(x|\underline{\theta}_j), \quad M \in \mathbb{N}.$$

Algoritmo EM (*Esperanza-Maximización*).

♠ La mayoría estamos familiarizados en hallar *estimadores de máxima verosimilitud*. Un método elegante y poderoso para hallarlos con variables latentes es el llamado *algoritmo EM (Esperanza-Maximización)*.

♠ Este algoritmo es determinista, mientras nosotros vamos a abordar los algoritmos estocásticos. Pero para poder llegar a ellos hay todo un camino por recorrer, comenzaremos a abrirnos brecha estudiando el algoritmo *EM*.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

Algoritmo EM (Dempster - Laird - Rubin 1977)

Dada una distribución conjunta $p(X, Z|\theta)$ sobre variables observables X y variables latentes Z , dado el parámetro θ . el objetivo es maximizar la función de verosimilitud $p(X|\theta)$ con respecto a θ

0. Se eligen valores iniciales para θ , denominado θ^{old} .
1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosimilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

EM mezcla de normales (Ejemplo)

Un modelo de mezclas finito de normales es el siguiente:

$$f(x) = \sum_{j=1}^M \pi_j \mathcal{N}(x | \mu_j, \sigma_j^2), \quad M \in \mathbb{N}$$

Dado un modelo de mezclas *Normales* nuestro fin es maximizar la funcion de verosimilitud respecto a los parámetros.

La función a maximizar es

$$\prod_{i=1}^N \sum_{j=1}^M \pi_j \mathcal{N}(x_i | \mu_j, \sigma_j^2), \quad N, M \in \mathbb{N}$$

EM mezcla de normales (Ejemplo)

Un modelo de mezclas finito de normales es el siguiente:

$$f(x) = \sum_{j=1}^M \pi_j \mathcal{N}(x | \mu_j, \sigma_j^2), \quad M \in \mathbb{N}$$

Dado un modelo de mezclas *Normales* nuestro fin es maximizar la función de verosimilitud respecto a los parámetros.

La función a maximizar es

$$\prod_{i=1}^N \sum_{j=1}^M \pi_j \mathcal{N}(x_i | \mu_j, \sigma_j^2), \quad N, M \in \mathbb{N}$$

Algoritmo EM (Mezcla de normales)

0. Inicializar las medias, desviaciones estándar

1. **Paso E:** Evaluar las responsabilidades usando los valores actuales

$$P(Z|X, \theta) = \gamma(Z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \sigma_j)}$$

Algoritmo EM (Mezcla de normales)

0. Inicializar las medias, desviaciones estándar
1. **Paso E:** Evaluar las responsabilidades usando los valores actuales

$$P(Z|X, \theta) = \gamma(Z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \sigma_j)}$$

EM mezcla de normales

2. **Paso M:** Estimamos nuevamente los parámetros usando las responsabilidades calculadas en el paso E

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$$

$$\sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

donde

$$N_k = \sum_{n=1}^N \gamma(z_{nk})$$

- ♣ Evaluar la log-verosimilitud , para ir monitoreando la convergencia del algoritmo

$$\ln p(X|\mu, \sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \sigma_k) \right\}$$

2. **Paso M:** Estimamos nuevamente los parámetros usando las responsabilidades calculadas en el paso *E*

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$$

$$\sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

donde

$$N_k = \sum_{n=1}^N \gamma(z_{nk})$$

- ♣ Evaluar la log-verosimilitud , para ir monitoreando la convergencia del algoritmo

$$\ln p(X|\mu, \sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \sigma_k) \right\}$$

2. **Paso M:** Estimamos nuevamente los parámetros usando las responsabilidades calculadas en el paso E

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$$

$$\sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

donde

$$N_k = \sum_{n=1}^N \gamma(z_{nk})$$

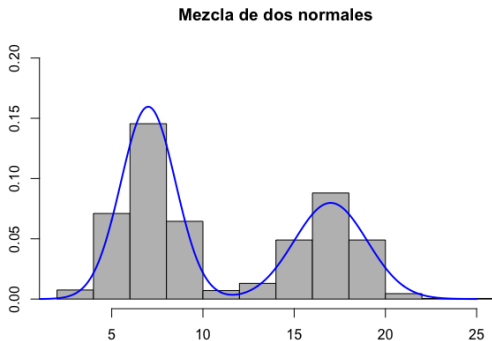
- ♣ Evaluar la log-verosimilitud , para ir monitoreando la convergencia del algoritmo

$$\ln p(X|\mu, \sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \sigma_k) \right\}$$

Ejemplo

Considere una mezcla de dos normales.

$$f(x) = 0,6\mathcal{N}(X|\mu_1, \sigma_1) + 0,4\mathcal{N}(X|\mu_2, \sigma_2)$$



Con $\mu_1 = 7$, $\mu_2 = 17$, $\sigma_1 = 1,5$ y $\sigma_2 = 2$

Nosotros vamos por inferencia Bayesianas
sobre mezclas infinitas...

$$f(x) = \sum_{j=1}^{\infty} \omega_j k_j(x|\underline{\theta}_j)$$

... O eso esperamos !!!

Nosotros vamos por inferencia Bayesianas
sobre mezclas infinitas...

$$f(x) = \sum_{j=1}^{\infty} \omega_j k_j(x|\underline{\theta}_j)$$

... O eso esperamos !!!

Nosotros vamos por inferencia Bayesianas
sobre mezclas infinitas...

$$f(x) = \sum_{j=1}^{\infty} \omega_j k_j(x|\underline{\theta}_j)$$

... O eso esperamos !!!

¡Muchas Gracias!

Nos vemos en la próxima travesía

El Dr. Yáñez está invitado...

Aunque se acaba de enterar