Tarea 5 Reconocimiento de Patrones

Yamil Ernesto Morfa

17 de marzo de 2021

B. Preguntas cortas

1. Tomamos el primer ejemplo del algoritmo MM, ver Recpat12n.pdf.

$$h_i(\theta \mid \theta^n) = 0{,}5\frac{(y_i-\theta)^2}{|y_i-\theta^n|} + 0{,}5\,|y_i-\theta^n|$$

$$f_i(\theta) = |y_i-\theta|$$

decimos que una función $g(\theta \mid \theta^n)$ mayoriza a $f(\theta)$ en θ^n si se cumple que:

$$\begin{gathered} f(\theta^n) = g(\theta^n \mid \theta^n) \\ f(\theta) \leq g(\theta \mid \theta^n) \quad \forall\,\theta \end{gathered}$$

Entonces tenemos que:

$$h_i(\theta^n \mid \theta^n) = 0{,}5\frac{(y_i-\theta^n)^2}{|y_i-\theta^n|} + 0{,}5\,|y_i-\theta^n| = (0{,}5+0{,}5)\,|y_i-\theta^n| = |y_i-\theta^n| = f(\theta^n)$$

$$h_i(\theta \mid \theta^n) = 0{,}5\frac{(y_i-\theta)^2}{|y_i-\theta^n|} + 0{,}5\,|y_i-\theta^n| = \frac{1}{|y_i-\theta^n|}\left\{\frac{(y_i-\theta)^2+|y_i-\theta^n|^2}{2}\right\} \geq \frac{1}{|y_i-\theta^n|}\left\{\sqrt{(y_i-\theta)^2\,|y_i-\theta^n|^2}\right\} = *$$

pues por la Desigualdad de las medias aritmética y geométrica nos dice que $\frac{a+b}{2} \geq \sqrt{ab}$

$$* = \sqrt{(y_i-\theta)^2} = |y_i-\theta| = f(\theta)$$

luego $h_i(\theta \mid \theta^n) \geq f(\theta) \quad \forall\,\theta$.

2. La desigualdad de Jensen establece que: Sean $f, g : \mathbb{R} \to \mathbb{R}$ con f convexa, entonces $E[f(g(X))] \geq f(E[g(X)])$. Ahora supongamos $X \sim P^1$ y $Y = g(X) = \frac{P^2(X)}{P^1(X)}$, sea $f(x) = -log(x)$ (convexa), entonces por la desigualdad de Jensen tenemos:

$$E_X[f(g(X))] \geq f(E_X[g(X)]) \Longrightarrow \begin{cases} \sum_i P^1(x_i)\,f(g(x_i)) \geq f\left(\sum_i P^1(x_i)\,g(x_i)\right) & \textit{caso discreto} \\ \int P^1(x)\,f(g(x))\,dx \geq f\left(\int P^1(x)\,g(x)\,dx\right) & \textit{caso continuo} \end{cases}$$

$$\Longrightarrow \begin{cases} -\sum_i P^1(x_i)\, log\left(\frac{P^2(x_i)}{P^1(x_i)}\right) \geq -log\left(\sum_i P^1(x_i)\,\frac{P^2(x_i)}{P^1(x_i)}\right) \\ -\int P^1(x)\, log\left(\frac{P^2(x)}{P^1(x)}\right) dx \geq -log\left(\int P^1(x)\,\frac{P^2(x)}{P^1(x)} dx\right) \end{cases}$$

$$\Longrightarrow \begin{cases} \sum_i P^1(x_i)\, log\left(\frac{P^1(x_i)}{P^2(x_i)}\right) \geq -log\left(\sum_i P^2(x_i)\right) = -\log 1 = 0 \\ \int P^1(x)\, log\left(\frac{P^1(x)}{P^2(x)}\right) dx \geq -log\left(\int P^2(x)\, dx\right) = -\log 1 = 0 \end{cases} \Longrightarrow d\left(P^1,\, P^2\right) \geq 0$$

C. Análisis de datos

1. Supongamos que tenemos un conjunto de datos $D = \{x_1, ..., x_N\}$ donde x_i es un vector de dimensión d. Supongamos ademas que estos son generados de manera aleatoria y $i.i.d$ por una densidad $p(x \mid \theta)$ definida como la mezcla finita de K densidades.

$$p(x \mid \theta) = \sum_{k=1}^{K} \alpha_k p_k(x \mid z_k, \theta_k)$$

con $z = (z_1, ..., z_K)$ con $z_k = \begin{cases} 0 \\ 1 \end{cases}$ y $\sum_{k=1}^{K} z_k = 1$, $\alpha_k = p_k(z_k)$ son los pesos de la mezcla, que representan que una variable x seleccionada aleatoriamente fue generada por la componente k, $\sum_{k=1}^{K} \alpha_k = 1$. El conjunto de parámetros de una modelo de K mezclas es

$$\theta = \{\alpha_1, ..., \alpha_K, \theta_1, ..., \theta_K\}$$

Podemos calcular

$$w_{ik} = p(z_{ik} = 1 \mid x_i, \theta) = \frac{\alpha_k p_k(x_i \mid z_k, \theta_k)}{\sum_{m=1}^{K} \alpha_m p_m(x_i \mid z_m, \theta_m)}, \qquad 1 \leq k \leq K, \quad 1 \leq i \leq N$$

Si asumimos que las densidades p_k son Gaussianas para cada componente con parámetros $\theta_k = \{\mu_k, \varSigma_k\}$

$$p_k(x \mid \mu_k, \varSigma_k) = \frac{1}{(2\pi)^{d/2} |\varSigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_k)^T \varSigma_k^{-1}(x - \mu_k)\right\}$$

El algoritmo EM es un algoritmo iterativo que comienza con una estimación aleatoria inicial de los parámetros θ y luego actualiza dichos parámetros hasta convergencia. Cada iteración consiste de dos pasos:

E-Step: Dado los valores actuales de θ calculamos W la matriz de pesos de $N \times K$ conformada por los valores $\{W\}_{ik} = w_{ik}$

M-Step: Utilizamos los pesos obtenidos para calcular los nuevos parámetros. Sea $N_k = \sum_{i=1}^{N} w_{ik}$ entonces:

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

$$\mu_k^{new} = \left(\frac{1}{N_k}\right) \sum_{i=1}^{N} w_{ik} x_i$$

$$\varSigma_k^{new} = \left(\frac{1}{N_k}\right) \sum_{i=1}^{N} w_{ik}(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

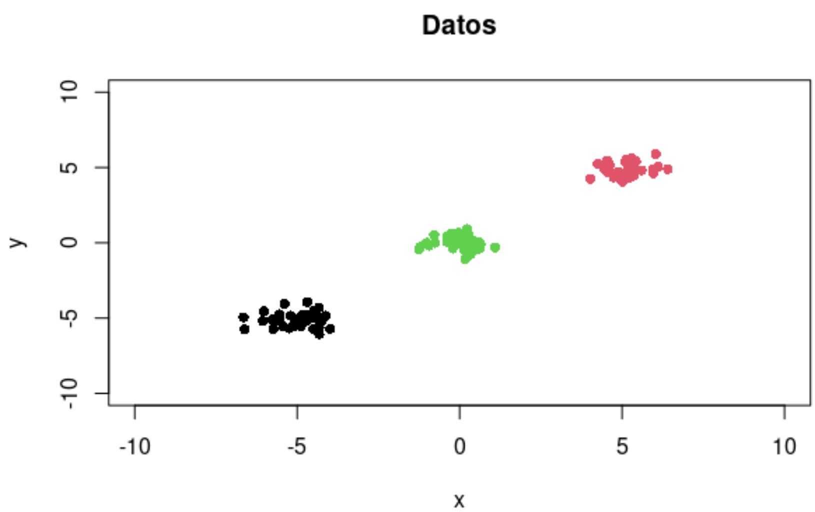
Convergencia:

La convergencia generalmente se detecta calculando el valor de la logverosimilitud después de cada iteración y deteniéndose cuando parece que no cambia de manera significativa de una iteración a la siguiente.

Ejemplo_1:

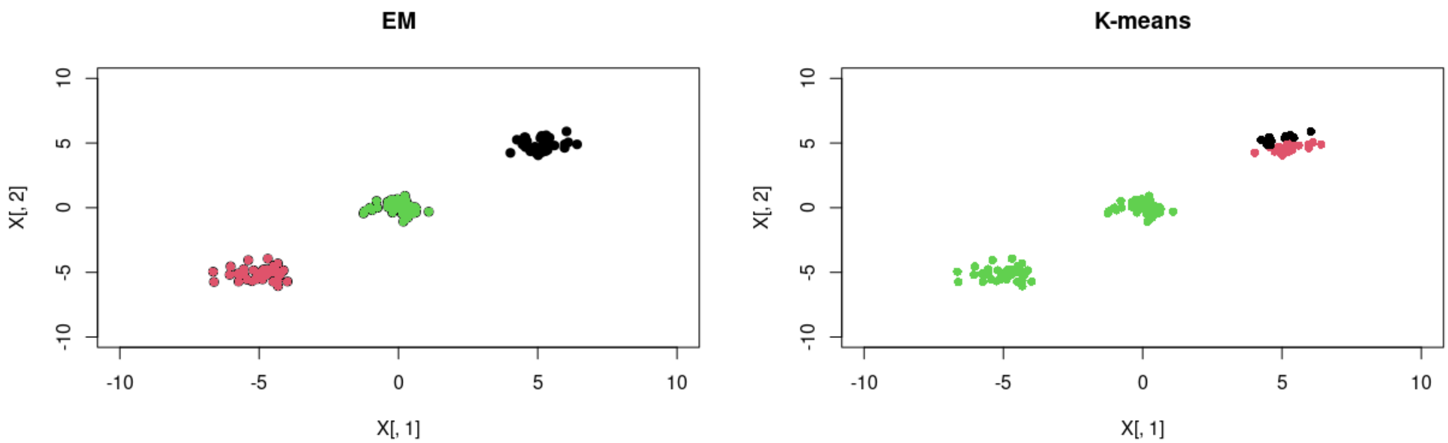
Para este ejemplo se toman datos de 3 normales bivariadas con $\mu_1 = (-5, -5)$, $\mu_2 = (0, 0)$, $\mu_3 = (5, 5)$ y $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0{,}5I$ como se muestran en la figura 1

Figura 1:



Como podemos ver este es un ejemplo en donde los grupos están muy bien separados. En consecuencia el algoritmo EM devuelve buenos resultados, sin embargo $K - means$ tiene problemas para agrupar bien estos datos. Esto se muestra en la Fig 2

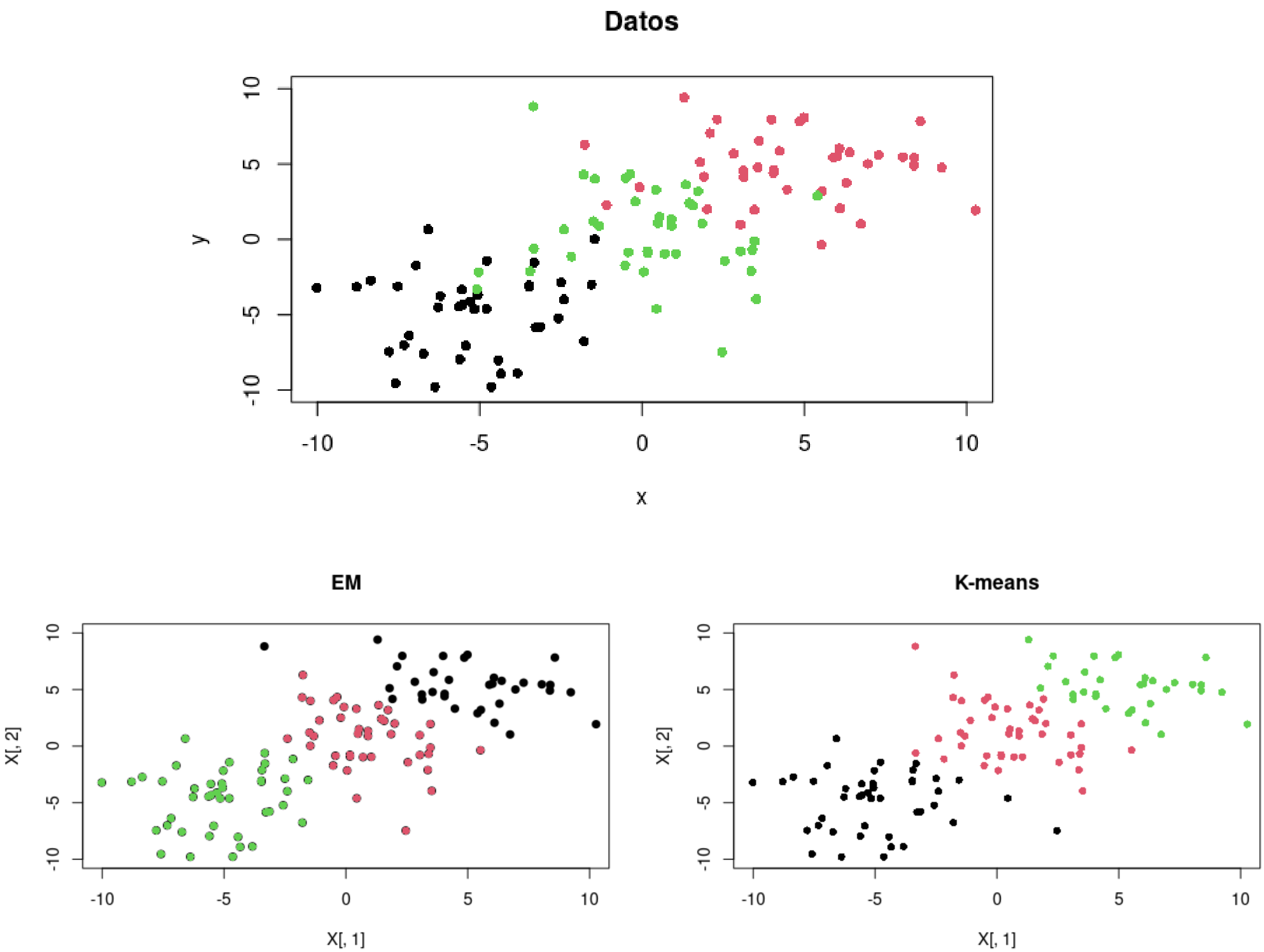
Figura 2:



Ejemplo_2:

Se Tomaron las normales bivariadas del ejemplo anterior solo que esta vez tomando $\Sigma_1 = \Sigma_2 = \Sigma_3 = 2{,}5I$. Los resultados obtenidos se muestran en la figura 3

Figura 3:



A pesar del buen comportamiento mostrado, EM tiene varios problemas en relación a la inicialización y a la convergencia.