HW7

↳ Hardik Prajapati (2678294168)

Q2. $C = \{1, 2\}$

Labeled samples = l

Unlabeled samples = $u = 1$

features = 1

$p(x \mid y = c, \underline{\theta}) = N(x \mid \mu_c, \sigma_c^2)$, $c = 1, 2$.

Priors & variances are constant.

Task :- Estimate μ_1, μ_2 using EM.

a) iteration :- t^{th}

$\boxed{\text{E step :- Compute best est of } H \text{ as } p(H \mid D, \underline{\theta}^{(t)})}$

$$\rightarrow p(H \mid D, \underline{\theta}^{(t)}) = \prod_{n=l+1}^{l+u} p(y_n \mid \underline{\underline{X}}_U, \underline{Y}_L, \underline{\underline{X}}_L, \underline{\theta}^{(t)})$$

$$= \prod_{n=l+1}^{l+u} p(y_n \mid \underline{x}_n, \underline{\theta}^{(t)})$$

As, $u = 1$

$$\Rightarrow p(H \mid D, \underline{\theta}^{(t)}) = p(y_n \mid \underline{x}_n, \underline{\theta}^{(t)})$$

$$= p(y_n = c_n \mid x_n, \underline{\theta}^{(t)}) \quad , c_n = 1, 2$$

$$= \gamma_{nc_n}^{(t)}$$

Now,

$$\gamma_{n c_n}^{(t)} = p\left(y_n = c_n \mid x_n, \underline{\theta}^{(t)}\right)$$

$$= \frac{p\left(x_n \mid y_n = c_n, \underline{\theta}^{(t)}\right) p\left(y_n = c_n \mid \underline{\theta}^{(t)}\right)}{\sum_{y_v=1}^{2} p\left(y_v = c_n \mid \underline{\theta}^{(t)}\right) p\left(x_n \mid y_v, \underline{\theta}^{(t)}\right)}$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma_{c_n}^2}} \exp\left(-\frac{1}{2} \frac{\left(x_n - \mu_{c_n}^{(t)}\right)^2}{\sigma_{c_n}^2}\right) \pi_{c_n}}{\left\{ (\pi_{c_1})\left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2} \frac{\left(x_n - \mu_1^{(t)}\right)^2}{\sigma_1^2}\right)\right) + (\pi_{c_2})\left(\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2} \frac{\left(x_n - \mu_2^{(t)}\right)^2}{\sigma_2^2}\right)\right) \right\}}$$

$$\Rightarrow \alpha_n^{(t)} = \frac{\pi_{c_1}}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2} \frac{\left(x_n - \mu_1^{(t)}\right)^2}{\sigma_1^2}\right) + \frac{\pi_{c_2}}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2} \frac{\left(x_n - \mu_2^{(t)}\right)^2}{\sigma_2^2}\right)$$

where π_{c_1} = Frequency of Labeled samples with class 1 (Priors)

π_{c_2} = Frequency of labeled samples with class 2 (Priors)

b. $p(D, H|\underline{\theta}) = p(H|D, \underline{\theta})\, p(D|\underline{\theta})$

$$p(D|\underline{\theta}) = \prod_{i=1}^{l} p(\underline{x}_i, y_i|\underline{\theta}) \prod_{n=l+1}^{l+u} p(\underline{x}_n|\underline{\theta})$$

$$= \left\{\prod_{i=1}^{l} p(x_i|y_i, \underline{\theta})\, p(y_i|\underline{\theta})\right\} \left\{\sum_{y=1}^{2} p(x_n|y, \underline{\theta})\, \pi_{c_y}\right\}$$

$$\Rightarrow p(D, H|\underline{\theta}) = y_{nc_n}\; p(D|\underline{\theta})$$

$$= \left\{\frac{p(x_n|y_n = c_n, \underline{\theta})\, p(y_n = c_n|\underline{\theta})}{\sum_{y_v=1}^{2} \underbrace{p(y_v|\underline{\theta})}_{\pi_{c_y}}\, p(x_n|y_v, \underline{\theta})}\right\} \cdot \left\{\sum_{y=1}^{2} p(x_n|y, \underline{\theta})\, \pi_{c_y}\right\}.$$

$$\left\{\prod_{i=1}^{l} p(x_i|y_i, \underline{\theta})\, \underbrace{p(y_i|\underline{\theta})}_{p(y_i = c_i)}\right\}$$

$$\Rightarrow p(D, H|\underline{\theta}) = p(x_n|y_n = c_n, \underline{\theta})\, \pi_{c_n} \prod_{i=1}^{l} p(x_i|y_i = c_i, \underline{\theta})\, \pi_{c_i}$$

where, $\pi_{c_i} = p(y_i = c_i|\underline{\theta}) = p(y_i = c_i)$

$\pi_{c_n} = p(y_n = c_n|\underline{\theta}) = p(y_n = c_n)$

$$\stackrel{c}{=} \ln p(D, H|\underline{\theta}) = \ln\{p(x_h|y_h = c_h, \underline{\theta})\} + \ln\{\pi_{c_h}\}$$

$$+ \sum_{i=1}^{l} \ln\{p(x_i|y_i = c_i, \underline{\theta})\} + \sum_{i=1}^{l} \ln\{\pi_{c_i}\}$$

→ Now, dropping terms that are not dependent on $\underline{\theta}$,

$$\Rightarrow \ln\{p(D, H|\underline{\theta})\} = \ln\{p(x_h|y_h = c_h, \underline{\theta})\} + \sum_{i=1}^{l} \ln\{p(x_i|y_i = c_i, \underline{\theta})\}$$

M Equation:-

$$\Rightarrow \underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}}\ E_{H|D, \underline{\theta}^{(t)}}\{\ln p(D, H|\underline{\theta})\}$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}} \sum_{c_h=1}^{2} y_{hc_h}^{(t)} \left\{ \ln\{p(x_h|y_h = c_h, \underline{\theta})\} + \sum_{i=1}^{l} \ln\{p(x_i|y_i = c_i, \underline{\theta})\} \right\}$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}} \left\{ \sum_{c_h=1}^{2} y_{hc_h}^{(t)} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma_{c_h}^2}} \exp\left(-\frac{1}{2} \frac{(x_h - \mu_{c_h})^2}{\sigma_{c_h}^2} \right) \right) \right] + \right.$$

(P.T.O)

$$\sum_{c_n=1}^{2} y_{nc_n}^{(t)} \left[\sum_{i=1}^{l} \ln\left(\frac{1}{\sqrt{2\pi\sigma_{c_i}^2}} \exp\left(-\frac{1}{2} \frac{(x_i - \mu_{c_i})^2}{\sigma_{c_i}^2} \right) \right) \right] \Bigg\}$$

(Dropping terms not Dependent on $\underline{\theta}$)

$$\Rightarrow \underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}} \Bigg\{ \sum_{c_n=1}^{2} y_{nc_n}^{(t)} \left[\frac{-1}{2} \frac{(x_n - \mu_{c_n})^2}{\sigma_{c_n}^2} \right] +$$

$$\left[\sum_{i=1}^{l} \left(\frac{-1}{2} \frac{(x_i - \mu_{c_i})^2}{\sigma_{c_i}^2} \right) \right] \left(y_{n1}^{(t)} + y_{n2}^{(t)} \right) \Bigg\}$$

$$\Rightarrow \underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}} \Bigg\{ \sum_{c_n=1}^{2} y_{nc_n}^{(t)} \left[-\frac{(x_n - \mu_{c_n})^2}{\sigma_{c_n}^2} \right] +$$

$$\sum_{i=1}^{l} \left[-\frac{(x_i - \mu_{c_i})^2}{\sigma_{c_i}^2} \right] \Bigg\}$$

d. l_1 = # Labeled Samples with $c_i = 1$
l_2 = # Labeled Samples with $c_i = 2$.
(from c)

$$\Rightarrow \underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}} \left\{ \sum_{c_n=1}^{2} y_{nc_n}^{(t)} \left[-\frac{(x_n - \mu_{c_n})^2}{\sigma_{c_n}^2} \right] + \sum_{i=1}^{l} \left[-\frac{(x_i - \mu_{c_i})^2}{\sigma_{c_i}^2} \right] \right\}$$

$$\Rightarrow \underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}} \left\{ y_{n1}^{(t)} \left[-\frac{(x_n - \mu_1)^2}{\sigma_1^2} \right] + y_{n2}^{(t)} \left[-\frac{(x_n - \mu_2)^2}{\sigma_2^2} \right] \right.$$

$$+ \sum_{i=1}^{l_1} [[c_i = 1]] \left\{ -\frac{(x_i - \mu_1)^2}{\sigma_1^2} \right\}$$

$$\left. + \sum_{i=1}^{l_2} [[c_i = 2]] \left\{ -\frac{(x_i - \mu_2)^2}{\sigma_2^2} \right\} \right\}$$

Now,

$$\frac{\partial}{\partial u_1} = y_{n1}^{(t)} \left[\frac{-2(x_n - \mu_1)(-1)}{\sigma_1^2} \right] + 0 + \sum_{i=1}^{l_1} [[c_i = 1]] \left\{ \frac{+2(x_i - \mu_1)}{\sigma_1^2} \right\} + 0.$$

$$\Rightarrow 0 = y_{n1}^{(t)} \frac{(x_n - \mu_1)}{\sigma_1^2} + \frac{1}{\sigma_1^2} \left\{ \sum_{i=1}^{l_1} [[c_i = 1]] x_i - l_1 \mu_1 \right\}$$

$$\Rightarrow 0 = \frac{y_{n1}^{(t)}(x_n) + \sum_{i=1}^{l_1} [[c_i = 1]] x_i}{\sigma_1^2} - \mu_1 \left\{ \frac{y_{n1}^{(t)} + l_1}{\sigma_1^2} \right\}$$

$$\Rightarrow \mu_1^{(t+1)} = \frac{y_{h1}^{(t)}(x_h) + \sum_{i=1}^{l_1} [[c_i = 1]] x_i}{\{y_{h1}^{(t)} + l_1\}}$$

Similarly,

$\frac{\partial}{\partial \upsilon_2} = 0$, we get

$$\mu_2^{(t+1)} = \frac{y_{h2}^{(t)}(x_h) + \sum_{i=1}^{l_2} [[c_i = 2]] x_i}{\{y_{h2}^{(t)} + l_2\}}$$

$$\therefore \underline{\theta}^{(t+1)} = \left[\frac{y_{h1}^{(t)}(x_h) + \sum_{i=1}^{l_1} [[c_i = 1]] x_i}{\{y_{h1}^{(t)} + l_1\}}, \frac{y_{h2}^{(t)}(x_h) + \sum_{i=1}^{l_2} [[c_i = 2]] x_i}{\{y_{h2}^{(t)} + l_2\}} \right]$$

e) i) $\gamma_{h1}(t) = 0.3486$ (Also computed on computer) (code file)

$\gamma_{h2}(t) = 0.6513$

$$\gamma_{h1}(t) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(3-1.5)^2}{1}\right) \cdot 0.5}{\frac{0.5}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(3-1.5)^2}{1}\right) + \frac{0.5}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(3-4)^2}{1}\right)}$$

$$= \frac{0.064788}{0.064788 + 0.120985}$$

$$= 0.3486$$

$\gamma_{h2}(t) = 0.651355$

ii) $$\mu_1(t+1) = \frac{(0.3486)(3) + (1+2)}{(0.3486) + 2}$$

$$= 1.72264327$$

$$\mu_2(t+1) = \frac{(0.6513)(3) + (4)}{(0.6513) + 1}$$

$$= 3.60558$$