HW - 13

Tamoghna
EE - 660

Q1 (a) The i^{th} iteration of EM, starting from :

$$p(\mathcal{H} \mid \mathcal{D}, \underline{\theta}^{(t)}) = p(y_n = c_n \mid \underline{x}_n, \underline{\theta}^{(t)}) = \gamma_{nc_n}^{(t)}$$

$$\gamma_{nc_n}^{(t)} = p(y_n = c_n \mid \underline{x}_n, \underline{\theta}^{(t)})$$

$$= \frac{p(\underline{x}_n \mid y_n = c_n, \underline{\theta}^{(t)}) \cdot p(y_n = c_n \mid \underline{\theta}^{(t)})}{p(\underline{x}_n \mid \underline{\theta}^{(t)})} \quad \text{--- (i)}$$

$$\because \; p(\underline{x}_n \mid \underline{\theta}^{(t)}) = \sum_{y_n=1}^{c_n} p(y_n \mid \underline{\theta}^{(t)}) \, p(\underline{x}_n \mid y_n, \underline{\theta}^{(t)})$$

Now,

$$p(\underline{x}_n \mid y_n = c_n, \underline{\theta}^{(t)}) = N(\underline{x}_n \mid \mu_{c_n}, \sigma_{c_n}^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma_{c_n}^2}} \exp\left\{ -\frac{1}{2\sigma_{c_n}^2} (x_n - \mu_{c_n}^{(t)})^2 \right\} \quad \text{--- (ii)}$$

Substituting (ii) in (i),

$$\gamma_{nc_n}^{(t)} = \frac{p(y_n = c_n \mid \underline{\theta}^{(t)})}{p(\underline{x}_n \mid \underline{\theta}^{(t)}) \sqrt{2\pi\sigma_{c_n}^2}} \exp\left\{ \frac{-1}{2\sigma_{c_n}^2} (x_n - \mu_{c_n}^{(t)})^2 \right\}$$

Hence, $\alpha_n^{(t)} = p(\underline{x}_n \mid \underline{\theta}^{(t)})$

and $\pi_{c_n} = p(y_n = c_n \mid \underline{\theta}^{(t)})$

(b) To derive an equation for $p(\mathcal{D}, H \mid \underline{\theta})$ for the Maximization formula,

$$p(\mathcal{D}, H \mid \underline{\theta}) = \underbrace{p(H \mid \mathcal{D}, \underline{\theta})}_{①} \underbrace{p(\mathcal{D} \mid \underline{\theta})}_{②}$$

① $\rightarrow$ $p(H \mid \mathcal{D}, \underline{\theta}) = p(y_h = c_h \mid \underline{x}_h, \underline{\theta})$

(given as the prior E step in the question)

Now,

$$p(y_h = c_h \mid \underline{x}_h, \underline{\theta}) = \frac{p(\underline{x}_h \mid y_h = c_h, \underline{\theta})\, p(y_h = c_h \mid \underline{\theta})}{p(\underline{x}_h \mid \underline{\theta})}$$

Also,

② $\rightarrow$ $p(\mathcal{D} \mid \underline{\theta}) = \prod_{i=1}^{l} p(x_i, y_i \mid \underline{\theta}) \cdot p(x_h \mid \underline{\theta})$

As there is only one unlabeled data.

Now, substituting these values of ① and ② back into the equation,

$$p(\mathcal{D}, H \mid \underline{\theta}) = \frac{p(\underline{x}_h \mid y_h = c_h, \underline{\theta})\, p(y_h = c_h \mid \underline{\theta})}{\cancel{p(\underline{x}_h \mid \underline{\theta})}} \times \prod_{i=1}^{l} p(x_i, y_i \mid \underline{\theta})\, \cancel{p(\underline{x}_h \mid \underline{\theta})}$$

$$= p(\underline{x}_h \mid y_h = c_h, \underline{\theta})\, p(y_h = c_h \mid \underline{\theta}) \times \prod_{i=1}^{l} p(x_i, y_i \mid \underline{\theta})$$

$$= p(\underline{x}_h \mid y_h = c_h, \underline{\theta})\, p(y_h = c_h \mid \underline{\theta}) \times \prod_{i=1}^{L} p(x_i \mid y_i = c_i ; \underline{\theta})\, p(y_i = c_i \mid \underline{\theta})$$

let $\pi_{c_i} = p(y_i = c_i \mid \underline{\theta})$

$\pi_{c_h} = p(y_h = c_h \mid \underline{\theta})$

$$\therefore p(\mathcal{D}, H \mid \underline{\theta}) = p(x_h \mid y_h = c_h, \underline{\theta})\, \pi_{c_h} \cdot \prod_{i=1}^{L} p(x_i \mid y_i = c_i, \underline{\theta})\, \pi_{c_i}$$

Hence Proved.

(c) From (b),

$$\ln [p(\mathcal{D}, H \mid \underline{\theta})] = \ln [p(\underline{x}_h \mid y_h = c_h, \underline{\theta})\, \pi_{c_h}] + \ln \left[\prod_{i=1}^{L} p(x_i \mid y_i = c_i, \underline{\theta}) \cdot \pi_{c_i} \right]$$

$$= \ln p(\underline{x}_h \mid y_h = c_h, \underline{\theta}) + \ln \pi_{c_h} + \ln \prod_{i=1}^{L} p(x_i \mid y_i = c_i, \underline{\theta}) + \ln \pi_{c_i}$$

By plugging normal densities in above, we get,

$$= \ln\left[\frac{1}{\sqrt{2\pi\sigma_{c_h}^2}} \exp\left\{-\frac{(x_h-\mu_{c_h})^2}{2\sigma_{c_h}^2}\right\}\right] + \ln \pi_{c_h}$$

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

$$+ \ln \pi_{c_i}$$

Dropping terms that are constants of $\underline{\theta}$,

$$\pi_{c_h} = p(y_h = c_h \mid \underline{\theta}) = p(y_h = c_h)$$
$$\pi_{c_i} = p(y_i = c_i \mid \underline{\theta}) = p(y_i = c_i)$$

So dropping them,

$$\ln[p(\mathcal{D}, H \mid \underline{\theta})] = \ln\left(\frac{1}{\sqrt{2\pi\sigma_{c_h}^2}}\right) + \left\{-\frac{(x_h-\mu_{c_h})^2}{2\sigma_{c_h}^2}\right\}$$
$$+ \sum_{i=1}^{l} \ln\left(\frac{1}{\sqrt{2\pi\sigma_{c_i}^2}}\right) + \sum_{i=1}^{l}\left\{-\frac{(x_i-\mu_{c_i})^2}{2\sigma_{c_i}^2}\right\}$$

$$\therefore \ln[p(\mathcal{D}, H \mid \underline{\theta})] =$$

$$= -\frac{1}{2}\left[\ln(2\pi\sigma_{c_h}^2) + \left\{-\frac{(x_h-\mu_{c_h})^2}{\sigma_{c_h}^2}\right\} + \sum_{i=1}^{l}\ln(2\pi\sigma_{c_i}^2)\right.$$
$$\left. + \sum_{i=1}^{l}\left\{-\frac{(x_i-\mu_{c_i})^2}{\sigma_{c_i}^2}\right\}\right]$$

Dropping constant multiplication factor of 2π,

$\therefore$

$$\ln p(\mathcal{D}, H | \underline{\theta}) = \frac{-(x_h - \mu_{ch})^2}{\sigma_{ch}^2} + \sum_{i=1}^{l} \frac{-(x_i - \mu_{ci})^2}{\sigma_{ci}^2}$$

Hence,

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

$$= \underset{\underline{\theta}}{\text{argmax}} \sum_{ch=1}^{2} \gamma_{hch}^{(t)} \ln p(\mathcal{D}, H | \theta)$$

$$= \underset{\underline{\theta}}{\text{argmax}} \sum_{ch=1}^{2} \gamma_{hch}^{(t)} \left[\frac{-(x_h - \mu_{ch})^2}{\sigma_{ch}^2} \right] + \sum_{i=1}^{L} \frac{-(x_i - \mu_{ci})^2}{\sigma_{ci}^2} \Bigg]$$

Hence proved.

(d) From (c),

$$\underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\text{argmax}} \left[\gamma_{hc_1}^{(t)} \left[\frac{-(x_h - \mu_1)^2}{\sigma_1^2} \right] + \gamma_{hc_2}^{(t)} \left[\frac{-(x_h - \mu_2)^2}{\sigma_2^2} \right] \right.$$

$$\left. + \sum_{i=1}^{l} \left[\frac{-(x_i - \mu_{ci})^2}{\sigma_{ci}^2} \right] \right] \quad \text{—— (1)}$$

Now taking derivative wrt μ_1,

$$\frac{\partial}{\partial \mu_1} \underline{\theta}^{(t+1)} = \frac{\partial}{\partial \mu_1} \left[\gamma_{hc_1}^{(t)} \left[\frac{-(x_h^2 + \mu_1^2 - 2x_h \mu_1)}{\sigma_1^2} \right] \right.$$

$$+ \gamma_{h_2}^{(t)} \left[\frac{-(x_h^2 + \mu_2^2 - 2x_h \mu_2)}{\sigma_2^2} \right]$$

$$+ \sum_{i=1}^{l_1} \left[\frac{-(x_i - \mu_1)^2}{\sigma_1^2} \right] + \sum_{i=1}^{l_2} \left[\frac{-(x_i - \mu_2)^2}{\sigma_2^2} \right]$$

$$\therefore \quad 0 = \delta_{h_1}^{(t)} \left[\frac{-2\mu_1 + 2x_h}{\sigma_1^2} \right] + \sum_{i=1}^{l_1} \left[\frac{-(2\mu_1 - 2x_i)}{\sigma_1^2} \right]$$

$$\therefore \quad \sum_{i=1}^{L_1} \mu_1 + \delta_{h_1}^{(t)} \mu_1 = \delta_{h_1}^{(t)} x_h + \sum_{i=1}^{l_1} x_i$$

$$\therefore \boxed{\mu_1^{(t+1)} = \frac{\gamma_{h_1}^{(t)} x_h + \sum_{i=1}^{l_1} x_i}{l_1 + \gamma_{h_1}^{(t)}}}$$

Similarly, if we derivate wrt μ_2 and solve by putting equal to 0, we will get,

$$\boxed{\mu_2^{(t+1)} = \frac{\gamma_{h_2}^{(t)} x_h + \sum_{i=1}^{l_2} x_i}{l_2 + \gamma_{h_2}^{(t)}}}$$

(e) (i) $$\gamma_{h_1}^{(t)} = \frac{\exp\left\{ -\frac{(3-1.5)^2}{2} \right\}}{\exp\left\{ -\frac{(3-1.5)^2}{2} \right\} + \exp\left\{ -\frac{(3-4)^2}{2} \right\}}$$

$$= 0.3486$$

$$\gamma_{h_2}^{(t)} = \frac{\exp\left\{ -\frac{(3-4)^2}{2} \right\}}{\exp\left\{ -\frac{(3-1.5)^2}{2} \right\} + \exp\left\{ -\frac{(3-4)^2}{2} \right\}}$$

$$= 0.6514$$

(ii) $$\mu_1^{(t+1)} = \frac{\gamma_{h_1}^{(t)} x_h + \sum_{i=1}^{L_1} x_i}{L_1 + \gamma_{h_1}^{(t)}}$$

$$= \frac{0.3486 \times 3 + 3}{2 + 0.3486}$$

$$= 1.7226$$

$$\mu_2^{(t+1)} = \frac{\gamma_{h_2}^{(t)} x_h + \sum_{i=1}^{l_2} x_i}{l_2 + \gamma_{h_2}^{(t)}}$$

$$= \frac{0.6514 \times 3 + 4}{1 + 0.6514}$$

$$= 3.6055$$