EE660 HW Week 13 Pin-Hsuan Lee 1139093874

Homework Problem – Semi-Supervised Learning (SSL)

1. Consider a 2-class semi-supervised learning problem in which there are l labeled samples and $u=1$ unlabeled sample. (For example, think of being given l labeled samples, and then acquiring unlabeled samples one at a time.) There is 1 feature, and each class is modeled as a Gaussian:

$$p\left(x \middle| y=c, \underline{\theta}\right) = N\left(x \middle| \mu_c, \sigma_c^{\,2}\right), \quad c=1,2$$

In the parts below, you will use EM to estimate the means μ_1 and μ_2 . You may assume the priors and variances are given constants. Generally the subscripts h and i will indicate unlabeled and labeled samples, respectively.

a) Consider the t^{th} iteration of EM. Derive the E step in terms of given quantities: that is, starting from

$$p\left(\mathcal{H} \middle| \mathcal{D}, \underline{\theta}^{(t)}\right) = p\left(y_h = c_h \middle| x_h, \underline{\theta}^{(t)}\right) = \gamma_{hc_h}^{(t)}, \quad c_h = 1,2$$

show that:

$$\gamma_{hc_h}^{(t)} = \frac{\pi_{c_h}}{\alpha_h^{(t)} \sqrt{2\pi\sigma_{c_h}^{\,2}}} \exp\left\{ -\frac{\left(x_h - \mu_{c_h}^{(t)}\right)^2}{2\sigma_{c_h}^{\,2}} \right\}$$

in which $\pi_{c_h} \triangleq p\left(y_h = c_h \middle| \underline{\theta}^{(t)}\right) = p\left(y_h = c_h\right)$. Also, find $\alpha_h^{(t)}$.

In parts (b)-(d), you will derive the M step formulas, also for the t^{th} iteration of EM.

b) First, show that

$$p\left(\mathcal{D}, \mathcal{H} \middle| \underline{\theta}\right) = p\left(x_h \middle| y_h = c_h, \underline{\theta}\right) \pi_{c_h} \prod_{i=1}^{l} p\left(x_i \middle| y_i = c_i, \underline{\theta}\right) \pi_{c_i}$$

in which $\pi_{c_i} = p\left(y_i = c_i \middle| \underline{\theta}\right) = p\left(y_i = c_i\right)$ and similarly for π_{c_h}.

c) Take $\ln p\left(\mathcal{D}, \mathcal{H} \middle| \underline{\theta}\right)$ from your result of (b), plug in for the normal densities, and drop any additive terms that are constants of $\underline{\theta}$. Then plug in to the M equation:

$$\begin{aligned} \underline{\theta}^{(t+1)} &= \underset{\underline{\theta}}{\arg\max}\ \mathsf{E}_{\mathcal{H}|\mathcal{D}, \underline{\theta}^{(t)}} \left\{ \ln p\left(\mathcal{D}, \mathcal{H} \middle| \underline{\theta}\right) \right\} \\ &= \underset{\underline{\theta}}{\arg\max} \sum_{c_h=1}^{2} \gamma_{hc_h}^{(t)} \ln p\left(\mathcal{D}, \mathcal{H} \middle| \underline{\theta}\right) \end{aligned}$$

and simplify to get:

$$\underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\arg\max} \left\{ \sum_{c_h=1}^{2} \gamma_{hc_h}^{(t)} \left[-\frac{\left(x_h - \mu_{c_h}\right)^2}{\sigma_{c_h}^{\,2}} \right] + \sum_{i=1}^{l} \left[-\frac{\left(x_i - \mu_{c_i}\right)^2}{\sigma_{c_i}^{\,2}} \right] \right\}$$

in which a constant multiplicative factor of $\frac{1}{2\pi}$ has been dropped. (**Hint:** you may find it useful to use $\gamma_{h1}^{(t)} + \gamma_{h2}^{(t)} = 1$.)

d) Re-write your result of part (c) to express it in terms of $\mu_1, \mu_2\ \sigma_1^2, \sigma_2^2$. (**Hint:** you might find it useful to use the indicator function.) Then solve for $\underline{\theta}^{(t+1)} = \left[\mu_1^{(t+1)}, \mu_2^{(t+1)}\right]^T$. (**Hint:** find the argmax by taking $\frac{\partial}{\partial \mu_1}$ and setting equal to 0; similarly for μ_2.) Let $l_1 =$ number of labeled samples with label $c_i = 1$, and $l_2 =$ umber of labeled samples with label $c_i = 2$. (Note that $\gamma_{hc_h}^{(t)}$ is constant of μ_1 and μ_2 because it used the (constant) estimates $\mu_1^{(t)}$ and $\mu_2^{(t)}$ from the E step.)

e) Given: $\pi_1 = \pi_2 = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$; data as follows:

labeled data $\left\{(x_i, y_i)\right\}_{i=1}^{l} = \left\{(1,1),(2,1),(4,2)\right\}$; unlabeled sample $x_h = 3$.

Suppose the values for $\underline{\theta}$ at the beginning of the t^{th} iteration of EM are: $\mu_1^{(t)} = 1.5$, $\mu_2^{(t)} = 4.0$. Note that you may solve this part (e) by hand, or use a computer to assist you (your choice).

(i) Calculate the responsibilities $\gamma_{h1}^{(t)}$ and $\gamma_{h2}^{(t)}$ from the E step (using part (a));

(ii) Calculate the new mean estimates $\mu_1^{(t+1)}$ and $\mu_2^{(t+1)}$ from the M step (using part (d) result).

1、

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

(a)

$$p(\mathcal{H}|\mathcal{D}, \underline{\theta}^{(t)}) = p(y_h = c_h | x_h, \underline{\theta}^{(t)}) = r_{hc_h}^{(t)}, \quad c_h = 1, 2$$

$$r_{hc_h}^{(t)} = p(y_h = c_h | x_h, \underline{\theta}^{(t)}) = \frac{p(x_h | y_h = c_h, \underline{\theta}^{(t)})\, p(y_h = c_h | \underline{\theta}^{(t)})}{\sum_{y_u=1}^{2} p(x_h | y_u, \underline{\theta}^{(t)})\, p(y_u | \underline{\theta}^{(t)})} \qquad (\text{Let } \pi_{y_u} = p(y_u))$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_{c_h}^2}} e^{-\frac{(x_h-\mu_{c_h})^2}{2\sigma_{c_h}^2}} \cdot \pi_{c_h}\right) \Big/ \underbrace{\sum_{y_u=1}^{2}\left(\pi_{y_u} \cdot \frac{1}{\sqrt{2\pi\sigma_{y_u}^2}} e^{-\frac{(x_h-\mu_{y_u})^2}{2\sigma_{y_u}^2}}\right)}_{\alpha_h^{(t)}}$$

$$= \frac{\pi_{c_h}}{\alpha_h^{(t)}\sqrt{2\pi\sigma_{c_h}^2}} e^{-\frac{(x_h-\mu_{c_h})^2}{2\sigma_{c_h}^2}} \quad \text{Q.E.D}$$

(b)

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

$$= p(y_h = c_h | x_h, \underline{\theta}) \cdot \left[\, p(x_h|\underline{\theta}) \cdot \prod_{i=1}^{l} p(x_i, y_i = c_i | \underline{\theta})\right]$$

$$= \frac{p(x_h | y_h = c_h, \underline{\theta})\, p(y_h = c_h | \underline{\theta})}{\sum_{y_u=1}^{2} p(x_h | y_u, \underline{\theta})\, p(y_u|\underline{\theta})} \cdot p(x_h|\underline{\theta}) \cdot \prod_{i=1}^{l} p(x_i | y_i = c_i, \underline{\theta}) \cdot p(y_i = c_i | \cancel{\underline{\theta}})$$

$$= \frac{p(x_h | y_h = c_h, \underline{\theta})\, \pi_{c_h}}{P(x_h, \underline{\theta}) / \cancel{p(\underline{\theta})}} \cdot \cancel{p(x_h|\underline{\theta})} \cdot \prod_{i=1}^{l} p(x_i | y_i = c_i, \underline{\theta})\, p(y_i = c_i)$$

$$= p(x_h | y_h = c_h, \underline{\theta}) \cdot \pi_{c_h} \cdot \prod_{i=1}^{l} p(x_i | y_i = c_i, \underline{\theta})\, \pi_{c_i} \quad \text{Q.E.D.}$$

(c) $\ln p(\mathcal{D},\mathcal{H}|\underline{\theta}) = \ln p(x_h|y_h=c_h,\underline{\theta}) + \cancel{\ln\pi_{c_h}} + \ln\left(\prod_{i=1}^{l} p(x_i|y_i=c_i,\underline{\theta})\right) + \cancel{\ln\pi_{c_i}}$

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

$$= \cancel{-\frac{1}{2}\ln(2\pi\sigma_{c_h}^2)} - \frac{(x_h-\mu_{c_h})^2}{\cancel{2}\sigma_{c_h}^2} \cancel{- \frac{l}{2}\ln(2\pi\sigma_{c_i}^2)} - \sum_{i=1}^{l}\frac{(x_i-\mu_{c_i})^2}{\cancel{2}\sigma_{c_i}^2}$$

$$\underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}}\ \mathbb{E}_{\mathcal{H}|\mathcal{D},\underline{\theta}^{(t)}}\left\{\ln p(\mathcal{D},\mathcal{H}|\underline{\theta})\right\}$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}} \left\{\sum_{H} p(\mathcal{H}|\mathcal{D},\underline{\theta}) \cdot \ln p(\mathcal{D},\mathcal{H}|\underline{\theta})\right\}$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}} \left\{\sum_{c_h=1}^{2} r_{hc_h}^{(t)}\left[-\frac{(x_h-\mu_{c_h})^2}{\sigma_{c_h}^2} - \sum_{i=1}^{l}\frac{(x_i-\mu_{c_i})^2}{\sigma_{c_i}^2}\right]\right\} \qquad * \ r_{h1}^{(t)} + r_{h2}^{(t)} = 1$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}} \left\{\sum_{c_h=1}^{2} r_{hc_h}^{(t)}\left[-\frac{(x_h-\mu_{c_h})^2}{\sigma_{c_h}^2}\right] - \sum_{i=1}^{l}\frac{(x_i-\mu_{c_i})^2}{\sigma_{c_i}^2}\right\} \qquad \text{Q.E.D}$$

(d)

$$\underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\operatorname{argmax}} \Big\{ \underbrace{r_{h1}^{(t)}\left[-\frac{(x_h-\mu_1)^2}{\sigma_1^2}\right] + r_{h2}^{(t)}\left[-\frac{(x_h-\mu_2)^2}{\sigma_2^2}\right] - \sum_{i=1}^{l}\frac{(x_i-\mu_1)^2}{\sigma_1^2}[\![c_i=1]\!] - \sum_{i=1}^{l}\frac{(x_i-\mu_2)^2}{\sigma_2^2}[\![c_i=2]\!]}_{f(x)}$$

$$\frac{\partial f(x)}{\partial \mu_1} = \frac{\cancel{2}r_{h1}^{(t)}(x_h-\mu_1)}{\cancel{\sigma_1^2}} + \frac{\cancel{2}}{\cancel{\sigma_1^2}}\sum_{i=1}^{l}(x_i-\mu_1)[\![c_i=1]\!] = 0$$

$$\Rightarrow r_{h1}^{(t)}x_h - r_{h1}^{(t)}\mu_1 - \mu_1 l_1 + \sum_{i=1}^{l} x_i [\![c_i=1]\!]$$

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

$$\frac{\partial f(x)}{\partial \mu_2} = \frac{\cancel{2}r_{h2}^{(t)}(x_{h2}-\mu_2)}{\cancel{\sigma_2^2}} + \frac{\cancel{2}}{\cancel{\sigma_2^2}}\sum_{i=1}^{l}(x_i-\mu_2)[\![c_i=2]\!] = 0$$

$$\Rightarrow r_{h2}^{(t)}x_h - r_{h2}^{(t)}\mu_2 - l_2\mu_2 + \sum_{i=1}^{l} x_i [\![c_i=2]\!]$$

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

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

(e)(i)

$$r_{h1}^{(t)} = \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_h-\mu_1)^2}{2\sigma_1^2}} \cdot \pi_1\right) \Big/ \left(\pi_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_h-\mu_1)^2}{2\sigma_1^2}} + \pi_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x_h-\mu_2)^2}{2\sigma_2^2}}\right)$$

$$\underset{x_h=3}{=} \frac{1}{\sqrt{2\pi\cdot 1}} e^{-\frac{(3-1.5)^2}{2\cdot 1}} \times \frac{1}{2} \Big/ \left(0.5 \times \frac{1}{\sqrt{2\pi\cdot 1}} e^{-\frac{(3-1.5)^2}{2\cdot 1}} + 0.5 \cdot \frac{1}{\sqrt{2\pi\cdot 1}} e^{-\frac{(3-4)^2}{2\cdot 1}}\right) \fallingdotseq 0.3486$$

$$r_{h2}^{(t)} = 1 - 0.3486 = 0.6514$$

(ii)

$$\underline{\theta}^{(t+1)} = \left[\frac{0.3486\cdot 3 + (1+2)}{2+0.3486},\ \frac{0.6514\cdot 3 + 4}{1+0.6514}\right]^T = [\underbrace{1.7226}_{\mu_1^{(t+1)}},\ \underbrace{3.6055}_{\mu_2^{(t+1)}}]^T$$