Dhruv Chakraborty
UID: 204-962-098

CM146: PSET 5

1)

(a) $L(h_t(x), \beta_t) = (e^{\beta t} - e^{-\beta t}) \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)] + e^{-\beta t} \sum_n w_t(n)$

$= (e^{\beta t} - e^{-\beta t}) \varepsilon_t + e^{-\beta t} \quad \because \sum_n w_t(n) = 1.$

$$\frac{\partial L}{\partial \beta_t} = (e^{\beta t} + e^{-\beta t}) \varepsilon_t - e^{-\beta t} = 0$$

$$\Rightarrow (e^{\beta t} + e^{-\beta t}) \varepsilon_t = e^{-\beta t}$$

$$\Rightarrow (e^{2\beta t} + 1) \varepsilon_t = 1$$

$$\Rightarrow e^{2\beta t} = \frac{1 - \varepsilon_t}{\varepsilon_t}$$

So $\beta_t^* = \frac{1}{2} \log\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$

(b) $\because$ training set is linearly separable

$\varepsilon_1 = \sum_n w_1(n) \mathbb{I}[y_n \neq h_1(x_n)] = 0$

So $\beta_t^* = \frac{1}{2} \log\left(\frac{1}{0}\right)$ or alternatively, we want

find β_t that minimizes $e^{-\beta t}$ which $\to 0$ as $\beta_t \to \infty$.

$\therefore \beta_t^* = \infty.$

2)

(a) The optimal clustering is $\{x_1 = 1, x_2 = 2\}, \{x_3 = 5\}, \{x_4 = 7\}$

Objective $J = (1-1.5)^2 + (2-1.5)^2 + (5-5)^2 + (7-7)^2$

$J = (0.5)^2 + (0.5)^2 = 0.5$

(b) If we start with the cluster assignment $\{x_1 = 1\}, \{x_2 = 2\}, \{x_3 = 5, x_4 = 7\}$ then we have $\mu_1 = 1, \mu_2 = 2, \mu_3 = 6$.

But then $J = (5-6)^2 + (7-6)^2 = 2 \not< 1/2 \to$ optimal

So this assignment is suboptimal.

Moreover, the algorithm ~~converges with~~ doesn't improve this as the clusters are stuck to their initial centroids as a form of local optima. $\therefore$ Lloyd's algorithm ~~is not~~ doesn't converge here.

3)

(a) $\nabla_{\mu_j} l(\theta) = \nabla_{\mu_j} \sum_n^i \gamma_{nj} \log \mathcal{N}(x_n | \mu_j, \Sigma_j)$

$$= \frac{\partial}{\partial \mu_j} \sum_n^i \gamma_{nj} \log \frac{1}{\sqrt{2\pi |\Sigma_j^{-1}|}} e^{\left(-\frac{1}{2}(x_n-\mu_j)^T \Sigma_j^{-1}(x_n-\mu_j)\right)}$$

$$= \frac{\partial}{\partial \mu_j} \sum_n^i \gamma_{nj} \left(-\frac{1}{2}(x_n-\mu_j)^T \Sigma_j^{-1}(x_n-\mu_j)\right)$$

$$= \sum_n^i \gamma_{nj} \left(-\Sigma_j^{-1}(x_n-\mu_j)\right)$$

$$= \sum_n^i \gamma_{nj} \Sigma_j^{-1} \mu_j - \sum_n \gamma_{nj} \Sigma_j^{-1} x_n$$

$$\nabla_{\mu_j} l(\theta) = \sum_n^i \gamma_{nj} \Sigma_j^{-1} (\mu_j - x_n)$$

(b) $\nabla_{\mu_j} l(\theta) = 0 \Rightarrow \cancel{\sum_n^i \gamma_n} \Sigma_j^{-1} \mu_j \sum_m^i \gamma_{nj} - \Sigma_j^{-1} \sum_n^i \gamma_{nj} x_n = 0$

$$\Rightarrow \mu_j \cdot \sum_n^i \gamma_{nj} = \sum_n^i \gamma_{nj} x_n$$

So $\mu_j = \frac{1}{\sum_n^i \gamma_{nj}} \sum_n^i \gamma_{nj} x_n$

(c) $w_1 = \frac{\sum_n^i \gamma_{1n}}{\sum_{ik} \sum_n \gamma_{kn}} = \frac{0.2+0.2+0.8+0.9+0.9}{(0.2+0.2+0.8+0.9+0.9)+(0.8+0.8+0.2+0.1+0.1)}$

$$= \frac{3}{5} = 0.6$$

$w_2 = \frac{\sum_n^i \gamma_{2n}}{\sum_{ik} \sum_n^i \gamma_{kn}} = \frac{2}{5} = 0.4$

$\mu_1 = \frac{1}{\sum_m \gamma_{1n}} \sum_n^i \gamma_{1n} x_n = \frac{(0.2)(5)+(0.2)(15)+(0.8)(25)(0.9)(30)+(0.9)(40)}{3}$

$$= \frac{87}{3} = 29$$

$\mu_2 = \frac{1}{\sum_n^i \gamma_{2n}} \sum_n^i \gamma_{2n} x_n = \frac{(0.8)(5)+(0.8)(15)+(0.2)(25)+(0.1)(30)+(0.1)(40)}{2}$

$$= \frac{28}{2} = 14$$