Christopher Ng
404775904

CS M146 Problem Set 5

① Adoboost

A) $(h_t^*(x), \beta_t^*) = \underset{(h_t(x),\beta_t)}{\operatorname{argmin}} \sum_n w_t(n) e^{-y_n \beta_t h_t(x_n)}$

$$= \underset{(h_t(x),\beta_t)}{\operatorname{argmin}} (e^{\beta_t} - e^{-\beta_t}) \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)] + e^{-\beta_t} \sum_n w_t(n)$$

$\epsilon_t = \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)]$ $\qquad \sum_n w_t(n) = 1$

$$(h_t^*(x), \beta_t^*) = \underset{(h_t(x),\beta_t)}{\operatorname{argmin}} (e^{\beta_t} - e^{-\beta_t}) \epsilon_t + e^{-\beta_t}(1)$$

$$= \underset{(h_t(x),\beta_t)}{\operatorname{argmin}} e^{\beta_t}\epsilon_t - e^{-\beta_t}\epsilon_t + e^{-\beta_t}$$

$$0 = \beta_t e^{\beta_t}\epsilon_t + \beta_t e^{-\beta_t}\epsilon_t - \beta_t e^{-\beta_t}$$

$$0 = \cancel{\beta_t} \cdot (e^{\beta_t}\epsilon_t + e^{-\beta_t}\epsilon_t - e^{-\beta_t})$$

$$0 = e^{\beta_t}\epsilon_t + e^{-\beta_t}\epsilon_t - e^{-\beta_t}$$

$$0 = e^{\beta_t}\epsilon_t + e^{-\beta_t}(\epsilon_t - 1)$$

$$-e^{\beta_t}\epsilon_t = e^{-\beta_t}(\epsilon_t - 1)$$

$$\log(e^{\beta_t}\epsilon_t) = \log(e^{-\beta_t}(\epsilon_t - 1))$$

$$\log(e^{\beta_t}) + \log(\epsilon_t) = \log(e^{-\beta_t}) + \log(\epsilon_t + 1)$$

$$\beta_t + \log(-\epsilon_t) = -\beta_t + \log(\epsilon_t + 1)$$

$$2\beta_t = \log(\epsilon_t + 1) - \log(-\epsilon_t)$$

$$2\beta_t = \log\left(\frac{\epsilon_t + 1}{-\epsilon_t}\right)$$

$$2\beta_t = \log\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

$$\boxed{\beta_t = \frac{1}{2}\log\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)}$$

$$\beta_1 = \frac{1}{2}\log\left(\frac{1-\epsilon_1}{\epsilon_1}\right)$$

B) $\epsilon_1 = \sum_n w_t(1)\, I(y_n \neq h_1(x_n))$

$h_1(x_n) = \text{sign}(w_1^T x_n + b)$

Because we are using hard-margin linear support vector machine as a base classifier, the SVM will perfectly classify the data. Thus,

$$\epsilon_1 = 0$$

$$\beta_1 = \frac{1}{2}\log\left(\frac{1-0}{0}\right)$$

$$\boxed{\beta_1 = \infty}$$

② k-means for single dimensional data

A) $k = 3$

$x_1 = 1$

$x_2 = 2$

$x_3 = 5$

$x_4 = 7$

The optimal clustering for this data would be $A(1) = 1$, $A(2) = 1$, $A(5) = 2$, $A(7) = 3$ where x_1 and x_2 are in cluster 1, x_3 is in cluster 2, and x_4 is in cluster 3

$$\mu_1 = 1.5, \quad \mu_2 = 5, \quad \mu_3 = 7$$

$$J(\{r_{nk}\},\{\mu_k\}) = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n - \mu_k\|_2^2$$

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

$$= (-0.5)^2 + (0.5)^2 + 0^2 + 0^2$$

$$= 0.25 + 0.25$$

$$\boxed{J(\{r_{nk}\},\{\mu_k\}) = 0.5}$$

B) Another clustering for this data could be:

$A(1) = 3$, $A(2) = 2$, $A(5) = 1$, $A(7) = 1$

where x_1 is in cluster 3, x_2 is in cluster 2, and x_3 and x_4 are in cluster 1.

$\mu_1 = \frac{5+7}{2} = 6$ $\mu_2 = 2$ $\mu_3 = 1$

$$J(\{r_{nk}\}, \{\mu_k\}) = 1(5-6)^2 + 1(7-6)^2 + (2-2)^2 + (1-1)^2$$
$$= (-1)^2 + (1)^2 + 0^2 + 0^2$$
$$= 1 + 1$$
$$J(\{r_{nk}\}, \{\mu_k\}) = 2$$

The value of the objective function of this clustering is suboptimal since $2 > 0.5$, which was the value of the obj. function of our other clustering in part A. The clustering is suboptimal because after μ_k is initialized, the data points x_n will be reassigned to the k clusters depending on which value of μ_k the point is closest to. Because the points x_n are closest to the cluster k that it is already assigned to, the reassignment will not have any changes on the next iteration, thus demonstrating that the value of the objective function will not improve.

(3) Gaussian Mixture Models

A) $$\ell(\theta) = \sum_n \log p(x_n, z_n)$$

$$= \sum_k \sum_n \gamma_{nk} \log \omega_k + \sum_k \left\{ \sum_n \gamma_{nk} \log N(x_n \mid \mu_k, \Sigma_k) \right\}$$

$$\nabla_{\mu_j} \ell(\theta) = 0 + \sum_n \frac{1}{\sum_k \gamma_{nk} N(x_n \mid \mu_k, \Sigma_k)} \gamma_{nk} N(x_n \mid \mu_k, \Sigma_k) \cdot \frac{(x_n - \mu_k)}{\Sigma_k^2}$$

$$\boxed{\nabla_{\mu_j} \ell(\theta) = \sum_n \frac{\gamma_{nk} N(x_n \mid \mu_k, \Sigma_k)}{\sum_k \gamma_{nk} N(x_n \mid \mu_k, \Sigma_n)} \cdot \frac{(x_n - \mu_k)}{\Sigma_k^2}}$$

B) $$\sum_n \frac{\gamma_{nk} N(x_n \mid \mu_k, \Sigma_n)}{\sum_k \gamma_{nk} N(x_n \mid \mu_k, \Sigma_n)} \cdot \frac{(x_n - \mu_k)}{\Sigma_k^2} = 0$$

$$\gamma_{nj} = \frac{\gamma_{nk} N(x_n ; \Sigma_k^2)}{\sum_k \gamma_{nk} N(x_n, \Sigma_k^2)}$$

$$\sum_n \gamma_{nj} \cdot \frac{(x_n - \mu_k)}{\Sigma_k^2} = 0$$

$$\frac{\sum_n \gamma_{nj} x_n}{\Sigma_k^2} - \frac{\sum_n \gamma_{nj} \mu_k}{\Sigma_k^2} = 0$$

$$\sum_n \gamma_{nj} x_n = \sum_n \gamma_{nj} \mu_k$$

$$\boxed{\mu_j = \frac{\sum_n \gamma_{nj} x_n}{\sum_n \gamma_{nj}}}$$

c) $K=2$, $N=5$, x_n, $n \in \{1, \dots, N\}$

$\{5, 15, 25, 30, 40\}$

$$w_k = \frac{\sum_n \gamma_{nk}}{\sum_k \sum_n \gamma_{nk}}$$

$$\mu_1 = \frac{1}{\sum_n \gamma_{n1}} \sum_n \gamma_{n1} x_n$$

$$= \frac{1}{0.2+0.2+0.8+0.9+0.9} \left[(0.2 \cdot 5) + (0.2 \cdot 15) + (0.8 \cdot 25) + (0.9 \cdot 30) + (0.9 \cdot 40) \right]$$

$$= \frac{1}{3} \left[1 + 3 + 20 + 27 + 36 \right]$$

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

$$\boxed{\mu_1 = 29}$$

$$\mu_2 = \frac{1}{\sum_n \gamma_{n2}} \sum_n \gamma_{n2} x_n$$

$$= \frac{1}{0.8+0.8+0.2+0.1+0.1} \left[(0.8 \cdot 5) + (0.8 \cdot 15) + (0.2 \cdot 25) + (0.1 \cdot 30) + (0.1 \cdot 40) \right]$$

$$= \frac{1}{2} \left[4 + 12 + 5 + 3 + 4 \right]$$

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

$$\boxed{\mu_2 = 14}$$

$$w_1 = \frac{\sum_n \gamma_{n1}}{\sum_k \sum_n \gamma_{nk}} = \frac{3}{3+2} = \frac{3}{5} \qquad \boxed{w_1 = \frac{3}{5}}$$

$$w_2 = \frac{\sum_n \gamma_{n2}}{\sum_k \sum_n \gamma_{nk}} = \frac{2}{3+2} = \frac{2}{5} \qquad \boxed{w_2 = \frac{2}{5}}$$