CS M146, Winter 2019
Problem Set 5: Boosting, Unsupervised Learning
Due March 16, 2019

Tian Ye
Collaborators: Derek Chu, Austin Guo

March 16, 2019

1 AdaBoost

(a) **Solution:**

$$(h_t^*(x), \beta_t^*) = \text{argmin}(e^{\beta_t} - e^{-\beta_t}) \sum_n w_t(n) \| [y_n \neq h_t(x_n)] \tag{1}$$

$$\Rightarrow 0 = (e^{\beta_t} - e^{-\beta_t})\epsilon_t - e^{-\beta_t^{\epsilon_t}} \tag{2}$$

$$0 = \epsilon_t e^{\beta_t} + \epsilon_t e^{-\beta_t} - e^{-\beta_t^{\epsilon_t}} \tag{3}$$

$$\epsilon_t e^{\beta_t} = e^{-\beta_t}(1 - \epsilon_t) \tag{4}$$

$$\beta_t + \ln \epsilon_t = -\beta_t + \ln(1 - \epsilon_t) \tag{5}$$

$$2\beta_t = \ln(1 - \epsilon_t) - \ln \epsilon_t \tag{6}$$

$$\beta_t = \frac{\ln(\frac{1-\epsilon_t}{\epsilon_t})}{2} \tag{7}$$

(b) **Solution:** $\beta_1 = \infty$ since this hard-margin linear SVM already classifies everything correctly as the training set is linearly seperable. Consequently we want β_1 to have all the weight as it already correctly classifies everything. Hence, we set it equal to infinity. Mathematically, $\epsilon_t = 0$ since nothing is ever misclassified; consequently this means that $\beta_t = \frac{\ln(\frac{1}{0})}{2} \Rightarrow \infty$.

2 K-means for single dimensional data

(a) **Solution:** Let μ_k represent teh kth cluster center. Using this notation, we are able to state that the optimal clustering is: $\mu_1 = 1.5 \mu_2 = 5, \mu_3 = 7$.

Here, x_1 and x_2 are in the first cluster, x_3 is in the second cluster, and x_4 is in the third cluster.

The cooresponding value of the objective is $0.5^2 + 0.5^2 + 0^2 + 0^2 = 0.5$

(b) **Solution:** If our initial cluster centers are $\mu_1 = 1, \mu_2 = 2, \mu_3 = 6$ and we use the same classification rules as in the previous section, we see that when we recalculate the means for Lloyd's algorithm, we get the same cluster centers and thus achieve convergence. Thus, this initial cluster assignment will not be improved because we have reached convergence on a local minimum. However, these assignments are suboptimal since the corresponding value of the objective is now $0^2 + 0^2 + 1^2 + 1^2 = 2$. Thus, the suboptimal initial clster assignment resulted in a convergence and getting a higher value for the objective, meaning that we did not converge to the local minimum.

3 Gaussian Mixture Models

(a) **Solution:**

$$\ell(\theta) = \sum_k \sum_n \gamma_{nk} \log \omega_k + \sum_k \{ \sum_n \gamma_{nk} \log N(x_n | \mu_k, \Sigma_k \} \tag{8}$$

$$\ell(\theta) = \sum_k \sum_n \gamma_{nk} \log \omega_k + \sum_k \sum_n \gamma_{nk} \log(\frac{1}{\sqrt{2\pi |\Sigma_k|}} e^{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)} \tag{9}$$

$$\ell(\theta) = \sum_k \sum_n \gamma_{nk} \log \omega_k + \sum_k \sum_n \gamma_{nk} \log(\frac{1}{\sqrt{2\pi |\Sigma_k|}} - \frac{1}{2}(x_n - \mu_i)^T \Sigma_k^{-1} (x_n - \mu_k)) \tag{10}$$

$$\nabla \mu_k \ell(\theta) = \sum_n \gamma_{nk} - \frac{1}{2} \Sigma_k^{-1} (x_n - \mu_k)(2)(-1) \tag{11}$$

$$\nabla \mu_k \ell(\theta) = \sum_n \gamma_{nk} \Sigma_k^{-1} (x_n - \mu_k) \tag{12}$$

(b) **Solution:**

$$0 = \sum_n \gamma_{nk}(x_n - \mu_k) \tag{13}$$

$$\sum_n \gamma_{nk} x_n = \sum_n \gamma_{nk}\mu_k \tag{14}$$

$$\mu_k = \frac{\sum_n \gamma_{nk} x_n}{\sum_n \gamma_{nk}} \tag{15}$$

(c) **Solution:** $x = \{5, 15, 25, 30, 40\}$

$\omega_k = \frac{\sum_n \gamma_{nk} x_n}{\sum_k \sum_n \gamma_{nk}}$, $\mu_k = \frac{1}{\sum_n \gamma_{nk}} \sum_n \gamma_{nk} x_n$ Therefore,

$$\omega_1 = \frac{(0.2 + 0.2 + 0.8 + 0.9 + 0.9)}{(0.2 + 0.2 + 0.8 + 0.9 + 0.9 + 0.8 + 0.8 + 0.2 + 0.1 + 0.1)} \tag{16}$$

$$\omega_1 = \tfrac{3}{5} \tag{17}$$

$$\omega_2 = \frac{(0.8 + 0.8 + 0.2 + 0.1 + 0.1)}{5} \tag{18}$$

$$\omega_2 = \tfrac{2}{5} \tag{19}$$

$$\mu_1 = \tfrac{1}{3}(0.2(5) + 0.2(15) + 0.8(25) + 0.9(30) + 0.9(40)) \tag{20}$$

$$\mu_1 = 29 \tag{21}$$

$$\mu_2 = \tfrac{1}{2}(0.8(5) + 0.8(15) + 0.2(25) + 0.1(30) + 0.1(40)) \tag{22}$$

$$\mu_2 = 14 \tag{23}$$