Shalin Shah
404930054

Problem Set 5

Problem 1. Adaboost @ t: $\arg\min_{(h_t^*(\vec{x}),\, \beta_t^*)} \sum_n w_t(n) e^{-y_n \beta_t h_t(\vec{x}_n)}$

a. We can expand this to the form of:

$(h_t^*(\vec{x}), \beta_t^*)$: $(e^{\beta_t} - e^{-\beta_t})\epsilon_t + e^{-\beta_t}$ since $\sum_n w_t(n) = 1$

Taking the partial: $\frac{\partial}{\partial \beta_t}\left[(e^{\beta_t} - e^{-\beta_t})\epsilon_t + e^{-\beta_t}\right] =$

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

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

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

$$e^{2\beta_t} = \frac{1-\epsilon_t}{\epsilon_t}$$

$$\boxed{\beta_t = \frac{1}{2}\log\left[\frac{1-\epsilon_t}{\epsilon_t}\right]}$$

b. Training set is linearly separable and hard margin SVM. $\epsilon_1 = 0$.

And our objective function goes to: $(h_1(\vec{x}), \beta_1) = (e^{\beta_1} - e^{-\beta_1}) \cdot 0 + e^{-\beta_1}$ $= e^{-\beta_1}$. The point is to minimize the objective function using $(h_t(\vec{x}), \beta_t)$, and β_1 has to be the value to minimize. So $\boxed{\beta_1 = \infty}$

Problem 2. K-means for single dimensional data.

a. $K = 3$, 4 data points: $x_1 = 1$, $x_2 = 2$, $x_3 = 5$, $x_4 = 7$.

Optimal clustering? Value of objective?

- If we draw a number line: 

We can clearly see that the optimal clustering (for the 1-D case) is $\boxed{(1,2), (5), (7).}$ The objective is thus:

$[1-1.5]^2 + [2-1.5]^2 = \boxed{0.5}$

b. One (but not us :)) might be tempted to assume Lloyd's algorithm will converge to the global minimum when $d=1$. ~~[illegible]~~ For our data we look to disprove this. When we set the means as 1.5, 5, and 7, we get $J = 0.5$ (as found in the previous part). If we set the means to be more offset, we will get a suboptimal solution. If $\mu_1 = 1.5$, $\mu_2 = 6$, $\mu_3 = 15$, we see that the points will be classified as cluster 1 or 2, but none as 3 and we get $J = (1.5-1)^2 + (1.5-2)^2 + (6-5)^2 + (6-7)^2 = 2.5$, which is greater than 1 and thus suboptimal. We're now at a local min, and thus the solution found is supoptimal.

Problem 3. We want to cluster data using a GMM with K mixture components.

Have to set $\vec{\theta} = \{w_k, \vec{\mu}_k, \vec{\Sigma}_k\}_{k=1}^{K}$ where w_k is weight.

MLE: $\ell(\vec{\theta}) = \sum_n \log p(\vec{x}_n, z_n) = \sum_k \sum_n \gamma_{nk} \log w_k + \sum_k \left\{ \sum_n \gamma_{nk} \log N(\vec{x}_n | \vec{\mu}_k, \vec{\Sigma}_k) \right\}$

a. Minimize ℓ: Find gradient $\nabla_{\vec{\mu}_j} \ell(\vec{\theta})$.

$$\frac{\partial \ell(\vec{\theta})}{\partial \mu_j} = 0 + \frac{\partial\left[\sum \sum \gamma_{nk} \log N(\vec{x}_n | \vec{\mu}_k, \vec{\Sigma}_k)\right]}{\partial \mu_j} =$$

$$= \frac{\partial\left[\sum_n \gamma_{nj} \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(\vec{x}_n - \mu_j)^2}{2\varepsilon_j}}\right)\right]}{\partial \mu_j} = \frac{\partial\left[\sum \gamma_{nj} \left\{\cancel{\log \sqrt{\varepsilon_j 2\pi}} + \log e^{-(\,)}\right\}\right]}{\partial \mu_j}$$

$$= \frac{\sum \gamma_{nj}\, \partial\left[\log e^{-\frac{(\vec{x}_n - \mu_j)^2}{2\varepsilon_j}}\right]}{\partial \mu_j} =$$

$$= \cancel{2} \sum_n \gamma_{nj} \frac{(\vec{x}_n - \mu_j)}{\cancel{2}\varepsilon_j} \Rightarrow \boxed{\nabla_{\vec{\mu}_j} \ell(\vec{\theta}) = \sum_n \frac{\gamma_{nj}(\vec{x}_n - \vec{\mu}_j)}{\varepsilon_j}}$$

b. Set this equal to 0 and solve for $\vec{\mu}_j$.

$$\sum_n \gamma_{nj} \frac{(\vec{x}_n - \vec{\mu}_j)}{\cancel{\sigma_j}} = 0 \rightarrow \sum_n \gamma_{nj}(\vec{x}_n) = \sum_n \gamma_{nj}(\vec{\mu}_j).$$

$$\boxed{\vec{\mu}_j = \frac{\sum_n \gamma_{nj}(\vec{x}_n)}{\sum_n \gamma_{nj}}}$$

This is exactly what we were to set it equal to. ✓✓

c. $K=2$, $N=5$. $\rightarrow w_1?\ w_2?\ \mu_1?\ \mu_2?$ Given Table 1.

$$w_1 = \frac{\sum \gamma_{n1}}{\sum \gamma_{nj}} = \frac{0.2+0.2+0.3+0.9+0.9}{5} = 0.6$$

$$w_2 = \frac{\sum \gamma_{n2}}{\sum \gamma_{nj}} = \longrightarrow = 0.4$$

$$\mu_1 = \frac{\sum_n \gamma_{n1} x_n}{3} = \frac{0.2\cdot 5 + 0.2\cdot 15 + 0.8\cdot 25 + 0.9\cdot 30 + 0.9\cdot 40}{3} = 29$$

$$\mu_2 = \frac{\sum \gamma_{n2} x_n}{2} = \frac{0.3\cdot 5 + 0.8\cdot 15 + 0.2\cdot 25 - 0.1\cdot 30 + 0.1\cdot 40}{2} = 14$$

$$\boxed{w_1 = 0.6,\ w_2 = 0.4,\ \mu_1 = 29,\ \mu_2 = 14}$$