CS M146 HW5

Saman Hashemipour
904 903 562

Problem 1a)

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

we are given $\sum_n w_t(n) = 1$ & we set $\epsilon_t = \sum_n w_t(n) \mathbb{I}[y_n \neq h_t(x_n)]$

So we can rewrite J as: $J(h_t(x), \beta_t) = (e^{\beta_t} - e^{-\beta_t})\epsilon_t + e^{-\beta_t}(1)$

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

Setting $\dfrac{dJ(h_t(x), \beta_t)}{d\beta_t} = 0$

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

$\Rightarrow -\beta_t + \ln(1-\epsilon_t) = \beta_t + \ln(\epsilon_t) \Rightarrow 2\beta_t = \ln(1-\epsilon_t) - \ln(\epsilon_t)$

$\Rightarrow \boxed{\beta_t = \ln\left(\dfrac{(1-\epsilon_t)}{\epsilon_t}\right) \cdot \dfrac{1}{2}}$

Problem 1b) Our predictor $h_1(x)$ for hard margin SVM is $h_1(x_n) = \text{sign}(w_t, x_n + b)$ and the weighted classification error ϵ_t is $\epsilon_1 = \sum_n w_1(n) \mathbb{I}[y_n \neq h_1(x_n)]$ since our data is linearly seperable $\epsilon_1 = 0$, SVM fits the data perfectly an no points get misclassified and when $\epsilon_t = 0$ our β_t approaches $+\infty$ because $\beta_t = \frac{1}{2}\ln\left(\frac{1-0^+}{0^+}\right) \Rightarrow \beta_t \to \infty$

$\boxed{\text{So our first iteration } \beta_1 \text{ would approach } \infty}$

Problem 2a) The optimum clustering for $K=3$ and $x = \{1, 2, 5, 7\}$ is:

$\boxed{\mu_1 = 1.5,\ \mu_2 = 5,\ \text{and } \mu_3 = 7}$ = centroids this is due to the fact that we assign points to the clusters based on the min distance which results in the clusters $(1,2), (5), \& (7)$ so we choose the updated centroids μ^* by $\mu_K = \dfrac{\sum_n r_{nK} x_n}{\sum_n r_{nK}} \Rightarrow$

$\mu_1^* = \frac{1+2}{2} = 1.5$
$\mu_2^* = 5/1 = 5$
$\mu_3^* = 7/1 = 7$

The centroids didn't change, so we have the optimum clustering.

The objective function is: $\sum_{n=1}^{N}\sum_{K=1}^{K} r_{nK}\|x_n - \mu_K\|_2^2 = (1-1.5)^2 + (2-1.5)^2 + (5-5)^2 + (7-7)^2 = \boxed{1/2}$

Problem 2b) Lloyds algoritm isn't guarenteed to converge to a gobal min. For cluster assignments $\{1\}, \{2\}, \{5, 7\}$ and $\mu_1 = 1$ $\mu_2 = 2$ $\mu_3 = 6$

In the update each point is assigned to the clusters as before bc the other centroids are farther

$x_1 : \|x_1 - u_j\|_2^2$ for $j=1 \Rightarrow 0$, $j=2 \Rightarrow 1$, $j=3 \Rightarrow 25$ $\rightarrow$ assign x_1 to cluster 1

$x_2 : \|x_2 - u_j\|_2^2$ for $j=1 \Rightarrow 1$, $j=2 \Rightarrow 0$, $j=3 \Rightarrow 16$ $\rightarrow$ assign x_2 to cluster 2

$x_3 : \|x_3 - u_j\|_2^2$ for $j=1 \Rightarrow 16$, $j=2 \Rightarrow 9$, $j=3 \Rightarrow 1$ $\rightarrow$ assign x_3 to cluster 3

$x_4 : \|x_4 - u_j\|_2^2$ for $j=1 \Rightarrow 36$, $j=2 \Rightarrow 16$, $j=3 \Rightarrow 1$ $\rightarrow$ assign x_4 to cluster 4

We see that this is not optimal b/c our cost function: $\sum_{n=1}^{N}\sum_{K=1}^{K} r_{nK}\|x_n - \mu_K\|_2^2 = 0^2 + 0^2 + 1^2 + 1^2 = 2$ which is greater than 1/2 from part A. Which makes part A's clustering more optimal than this one.