CM146, Fall 2017
Problem Set 5: Boosting, Unsupervised Learning
Due Dec 7, 2017

Author: Shivraj Gill

1 Problem 1

(a) Problem 1a

Solution: Solution to problem 1a

Our objective function is:

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

Because $\epsilon_t = \sum_n w_t(n)\mathbb{I}[y_n \neq h_t(x_n)]$ and $\sum_n w_t(n) = 1$, we can write

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

We now take the derivative of J w.r.t β_t and solve for β_t.

$$\begin{aligned}
\frac{\partial J(h_t(x), \beta_t)}{\partial \beta_t} &= \epsilon_t e^{\beta_t} + \epsilon_t e^{-\beta_t} - e^{-\beta_t} = 0 \\
&\Rightarrow \epsilon_t e_t^{\beta} = e^{-\beta_t}(1 - \epsilon_t) \\
&= log(\epsilon_t) + \beta_t = log(1 - \epsilon_t) - \beta_t \\
&= 2\beta_t = log(\frac{1 - \epsilon_t}{\epsilon_t}) \\
&\Rightarrow \beta_t = \frac{1}{2} log(\frac{1 - \epsilon_t}{\epsilon_t})
\end{aligned}$$

(b) Problem 1b

Solution: Solution to problem 1b

Initially, our algorithm sets our weights $w_1(n) = \frac{1}{N}$ for each sample. Next we train our weak classifier, $h_1(x)$, to minimize

$$\epsilon_1 = \sum_n w_1(n)\mathbb{I}[y_n \neq h_1(x_n)]$$
$$\beta_1 = \frac{1}{2}log(\frac{1-\epsilon_1}{\epsilon_1})$$

Because our data is linearly separable, our linear hard-margin SVM should get a minimum error of $\epsilon_1 = 0$.
This implies $\beta_1 = +\infty$.

2 Problem 2

(a) Problem 2a

Solution: Solution to problem 2a

K = 3, $x_1 = 1$, $x_2 = 2$, $x_3 = 5$, $x_4 = 7$
The optimal clustering is $\mu_1 = 1.5$, $\mu_2 = 5$, $\mu_3 = 7$,
where x_1 x_2 belong to cluster 1, x_3 belongs to cluster 2, and x_4 belongs to cluster 3.
The value of the objective function, J, is:

$$J = (1 - 1.5)^2 + (2 - 1.5)^2 + (5 - 5)^2 + (7 - 7)^2 = 0.5$$

(b) Problem 2b

Solution: Solution to problem 2a

A possible suboptimal assignment could be:
$\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 6$,
where x_1 is in cluster 1, x_2 is in cluster 2, and x_3, x_4 are in cluster 3
The value of the objective function, J, is

$$J = (1 - 1)^2 + (2 - 2)^2 + (5 - 6)^2 + (7 - 6)^2 = 2$$

Thus, this assignment is suboptimal since the value of our objective is less than the one in part a), i.e $2 < 0.5$.
Fixing our cluster assignments, we can calculate the new centroids: $\mu_1 = \frac{x_1}{1} = 1$, $\mu_2 = \frac{x_2}{1} = 2$
$\mu_3 = \frac{x_3 + x_4}{2} = 6$.

Because our centroid values are unchanged, the cluster assignments will remain the same, and so Lloyd's algorithm will stop looping and we will finish with this suboptimal solution.

3 Problem 3

(a) Problem 3a

Solution: Solution to problem 3a

We are missing the transition probabilities:

$$q_{21} = P(q_{t+1} = 2|q_t = 1)$$
$$q_{22} = P(q_{t+1} = 2|q_t = 2)$$

Because $\sum_k q_{k1} = 1$ and $\sum_k q_{k2} = 1$, $q_{21}, q_{22} = 0$. Since the logic is the same for both transition probabilities, we'll only show the derivation for q_{21}.

$$\begin{aligned}\sum_k q_{k1} = 1 &= q_{11} + q_{21} \\ &= 1 + q_{21} \\ &\Rightarrow q_{21} = 0\end{aligned}$$

We are also missing the output probabilities:

$$e_1(B) = P(O_t = B|q_t = 1)$$
$$e_2(A) = P(O_t = A|q_t = 2)$$

Using the constraint $\sum_b e_i(b) = 1$, we can solve for $e_1(B)$ and $e_2(A)$.

$$\begin{aligned}\sum_b e_1(b) = 1 &= e_1(A) + e_1(B) \\ &= 0.99 + e_1(B) \\ &\Rightarrow e_1(B) = 0.01\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_b e_2(b) = 1 &= e_2(A) + e_2(B) \\ &= e_2(A) + 0.49 \\ &\Rightarrow e_2(A) = 0.51\end{aligned}$$

Thus we get, $q_{21}, q_{22} = 0$ and $e_2(A) = 0.51$, $e_1(B) = 0.01$.

(b) Problem 3b

Solution: Solution to problem 3b

We will solve $P(O_1 = A)$ and $P(O_2 = B)$ and choose the symbol with the highest probability.

$$\begin{aligned}
P(O_1 = A) &= \sum_k P(O_1 = A, q_1 = k) \\
&= P(O_1 = A|q_1 = 1)P(q_1 = 1) + P(O_1 = A|q_1 = 2)P(q_1 = 2) \\
&= e_1(A)\pi_1 + e_2(A)\pi_2 \\
&= 0.99 * 0.49 + 0.49 * 0.51 \\
&= 0.735
\end{aligned}$$

Likewise,

$$\begin{aligned}
P(O_1 = B) &= \sum_k P(O_1 = B, q_1 = k) \\
&= P(O_1 = B|q_1 = 1)P(q_1 = 1) + P(O_1 = B|q_1 = 2)P(q_1 = 2) \\
&= e_1(B)\pi_1 + e_2(B)\pi_2 \\
&= 0.01 * 0.49 + 0.51 * 0.51 \\
&= 0.265
\end{aligned}$$

Since $P(O_1 = A) > P(O_1 = B)$, A is the most frequent symbol to appear in the first position of sequences of this HMM model.

(c) Problem 3b

Solution: Solution to problem 3b

We want to maximize the joint probability $P(O_{1:3}, q_{1:3})$. Because $q_{11}, q_{12} = 1$, our states for the second and third symbol will be 1. Then, the probability $P(O_t = A)$ and $P(O_t = B)$ for t = 1, 2 can be computed:

$$\begin{aligned}
P(O_t = A) = P(A|q_t = 1)P(q_t = 1) = 0.99 * 1 = 0.99 \\
P(O_t = B) = P(B|q_t = 1)P(q_t = 1) = 0.01 * 1 = 0.01
\end{aligned}$$

Thus, the most frequent symbol for t = 2, 3 is AA since $P(O_t = A)$ ¿ $P(O_t = B)$. And from part b) we know that A is the most frequent symbol for t = 1, so AAA is the most probable path.