CS 6780: Advanced Machine Learning (Submitted on: 4/18/2019)

Submission Assignment #4

Instructor: Thorsten Joachims *Name:* Molly Ingram, Julien Neves, *Netid:* msi34, jmn252

Problem 1

(a) To compute the probability of observing $(O_1, O_2, O3) = (0, 1, 0)$ we must compute the probability of observing this in all possible state sequences, shown in the table below. Summing these probabilities gives $P((O_1, O_2, O3) = (0, 1, 0)) = 0.12038$.

State Sequence	Initial Prob X Transition Probs	Probability
AAA	$P(A)P(0\|A)P(A\|A)P(1\|A)P(A\|A)P(0\|A)$	0.12
AAB	$P(A)P(0\|A)P(A\|A)P(1\|A)P(B\|A)P(0\|B)$	0.00016
ABB	$P(A)P(0\|A)P(B\|A)P(1\|B)P(B\|B)P(0\|B)$	0.000057
ABA	$P(A)P(0\|A)P(B\|A)P(1\|B)P(A\|B)P(0\|A)$	0.000071
BAA	$P(B)P(0\|B)P(A\|B)P(1\|A)P(A\|A)P(0\|A)$	0.0000016
BAB	$P(B)P(0\|B)P(A\|B)P(1\|A)P(B\|A)P(0\|B)$	0.00000002
BBA	$P(B)P(0\|B)P(B\|B)P(1\|B)P(A\|B)P(0\|A)$	0.0000071
BBB	$P(B)P(0\|B)P(B\|B)P(1\|B)P(B\|B)P(0\|B)$	0.000088

(b) Using the Viterbi algorithm to compute the most likely sequence is shown in the table below. They states chosen by the algorithm are highlight in red.

State	$P(y)P(0\|y)$	$P(y\|y_1)P(1\|y)$	$P(y\|y_2)P(0\|y)$
A	0.79	0.198	0.79
B	0.001	0.009	0.001

Problem 2

(a) Consider $X = (0, 1, 1)$, which generates $H = (n_1 = 0, h_2 = 1, h_3 = 1)$, and $y^* = h^*(x_t) = (1, 0, 1)$ so $d = 3$. For the first iteration through the algorithm, we make a mistake and reduce $V_2 = h_2, h_3$. The second iteration we also make a mistake and we get $V_3 = h_3$. The final iteration ends with h_3 as the only consistent hypothesis after making $d-1$ mistakes. However, it's possible to end the algorithm with no consistent hypotheses. If we had $d = 2$ and removed the 3rd example, we would make 2 mistakes and end with V as the empty set.

(b) Consider the Halving algorithm presented in lecture with the one additional rule that in the event of no simple majority, the algorithm chooses the lowest expert prediction.

Now consider $X = (0, 1, 1)$ and $h^* = (1, 1, 1)$. The table below lists the hypothesis class and when we arrive at the perfect expert. It is clear from the table that the algorithm makes exactly $log_2|H| = 3$ mistakes.

H	Iter. 1	Iter. 2	Iter. 3	End
(0,1,1)	X			
(0,0,1)	X			
(0,1,0)	X			
(0,0,0)	X			
(1,1,1)				Perfect
(1,0,1)		X		
(1,0,0)		X		
(1,1,0)			X	

(c)

Problem 3

(a) First, we compute the $P(x_i \mid p^{(k)})$ using the following formula

$$P(x \mid p^{(k)}) = \Pi_{d=1}^{D} (p_d^{(k)})^{x_d} (1 - p_d^{(k)})^{(1-x_d)}$$

and we report the results in Table 1

x_i	$P(x_i \mid (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}))$	$P(x_i \mid (0,0,0))$
$(0,1,0)$	$\frac{1}{8}$	0
$(0,1,1)$	$\frac{1}{8}$	0
$(0,0,0)$	$\frac{1}{8}$	1

Table 1: $P(x_i \mid p^k)$ at time $t = 0$

By definition, we have that

$$P(x_i \mid z_i^{(k)}, \pi, \boldsymbol{p}) = P(x_i \mid p^k)$$

This proprety simplifies the expression $\eta(z_i^{(k)})$ with the following formula

$$\begin{aligned}
\eta(z_i^{(k)}) &= \frac{\pi_k P(x_i \mid z_i^k, \pi, \boldsymbol{p})}{\sum_{k\prime} \pi_{k\prime} P(x_i \mid z_i^{k\prime}, \pi, \boldsymbol{p})} \\
&= \frac{\pi_k P(x_i \mid p^k)}{\sum_{k\prime} \pi_{k\prime} P(x_i \mid p^{k\prime})}
\end{aligned}$$

Therefore, using Table 1, we can compute easily the $\eta(z_i^{(k)})$. The results are reported in Table 2.

$i \backslash k$	$k = 1$	$k = 2$
$i = 1$	1	0
$i = 2$	1	0
$i = 3$	$\frac{1}{9}$	$\frac{8}{9}$

Table 2: $\eta(z_i^{(k)})$ at time $t = 0$

(b)

First, we can sum the $\eta(z_i^{(k)})$ over the i, to get N_k, i.e.,

$$N_1 = 1 + 1 + \frac{1}{9} = \frac{19}{9} \text{ and } N_2 = \frac{8}{9}$$

Using, x_i, $\eta(z_i^{(k)})$, and N_k, it is straightforward to compute $p_{t=1}^{(1)}$ and $p_{t=1}^{(2)}$, i.e.,

$$\begin{aligned}
p_{t=1}^{(1)} &= \frac{\sum_{i=1}^{N} \eta(z_i^{(1)}) x_i}{N_1} \\
&= \frac{1 \cdot (0,1,0) + 1 \cdot (0,1,1) + \frac{1}{9} \cdot (0,0,0)}{\frac{19}{9}} \\
p_{t=1}^{(1)} &= \left(0, \frac{18}{19}, \frac{9}{19}\right)
\end{aligned}$$

and

$$\begin{aligned}
p_{t=1}^{(2)} &= \frac{\sum_{i=1}^{N} \eta(z_i^{(2)}) x_i}{N_2} \\
&= \frac{0 \cdot (0,1,0) + 0 \cdot (0,1,1) + \frac{8}{9} \cdot (0,0,0)}{\frac{8}{9}} \\
p_{t=1}^{(2)} &= (0,0,0)
\end{aligned}$$

Therefore, $\boldsymbol{p}_{t=1}$ is given by

$$\boldsymbol{p}_{t=1} = \left\{ \left(0, \frac{18}{19}, \frac{9}{19}\right), (0,0,0) \right\}$$

(c)

Since we have computed the N_k in part (b), it is even easier now to update π. In fact, we have that

$$\begin{aligned}\pi_{k=1,t=1} &= \frac{N_1}{N_1+N_2} \\ &= \frac{\frac{19}{9}}{\frac{19}{9}+\frac{8}{9}} \\ \pi_{k=1,t=1} &= \frac{19}{27}\end{aligned}$$

and

$$\begin{aligned}\pi_{k=1,t=1} &= \frac{N_1}{N_1+N_2} \\ &= \frac{\frac{8}{9}}{\frac{19}{9}+\frac{8}{9}} \\ \pi_{k=2,t=1} &= \frac{8}{27}\end{aligned}$$

Combining both solution, we have the following

$$\pi_{t=1} = \left\{ \frac{19}{27}, \frac{8}{27} \right\}$$

(d)

In order, for us to be able to compute the loglikelihood, we first need $P(x_i \mid \boldsymbol{p}, \pi) = \sum_{k=1}^K \pi_k P(x_i \mid p^{(k)})$. To do this, we first reproduce Table 1, and add a column for $\sum_{k=1}^K \pi_k P(x_i \mid p^{(k)})$ where $\pi_{t=0} = \left\{\frac{1}{2}, \frac{1}{2}\right\}$

x_i	$P(x_i \mid (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}))$	$P(x_i \mid (0,0,0))$	$\sum_{k=1}^K \pi_k P(x_i \mid p^{(k)})$
$(0,1,0)$	$\frac{1}{8}$	0	$\frac{1}{2}\cdot\frac{1}{8}+\frac{1}{2}\cdot 0 = \frac{1}{16}$
$(0,1,1)$	$\frac{1}{8}$	0	$\frac{1}{2}\cdot\frac{1}{8}+\frac{1}{2}\cdot 0 = \frac{1}{16}$
$(0,0,0)$	$\frac{1}{8}$	1	$\frac{1}{2}\cdot\frac{1}{8}+\frac{1}{2}\cdot 1 = \frac{9}{16}$

Table 3: $P(x_i \mid p^k)$ at time $t = 0$

Using Table 3, it is easy to compute $L(X \mid \boldsymbol{p}_{t=0}, \pi_{t=0})$, i.e.,

$$\begin{aligned}L(X \mid \boldsymbol{p}_{t=0}, \pi_{t=0}) &= \sum_{i=1}^N \log P(x_i \mid \boldsymbol{p}_{t=0}, \pi_{t=0}) \\ &= \log\left(\frac{1}{16}\right) + \log\left(\frac{1}{16}\right) + \log\left(\frac{9}{16}\right) \\ L(X \mid \boldsymbol{p}_{t=0}, \pi_{t=0}) &\approx -6.12\end{aligned}$$

For $L(X \mid \boldsymbol{p}_{t=1}, \pi_{t=1})$, we repeat exactly the previous steps.

x_i	$P(x_i \mid (0, \frac{18}{19}, \frac{9}{19}))$	$P(x_i \mid (0,0,0))$	$\sum_{k=1}^K \pi_k P(x_i \mid p^{(k)})$
$(0,1,0)$	$\frac{18}{19}\cdot\frac{10}{19} = \frac{180}{361}$	0	$\frac{19}{27}\cdot\frac{180}{361}+\frac{8}{27}\cdot 0 = \frac{20}{57}$
$(0,1,1)$	$\frac{18}{19}\cdot\frac{9}{19} = \frac{162}{361}$	0	$\frac{19}{27}\cdot\frac{162}{361}+\frac{8}{27}\cdot 0 = \frac{6}{19}$
$(0,0,0)$	$\frac{1}{19}\cdot\frac{10}{19} = \frac{10}{361}$	1	$\frac{19}{27}\cdot\frac{10}{361}+\frac{8}{27}\cdot 1 = \frac{6}{19}$

Table 4: $P(x_i \mid p^k)$ at time $t = 0$

Using Table 4, we compute $L(X \mid \boldsymbol{p}_{t=1}, \pi_{t=1})$ in the following way

$$\begin{aligned}
L(X \mid \boldsymbol{p}_{t=1}, \pi_{t=1}) &= \sum_{i=1}^{N} \log P(x_i \mid \boldsymbol{p}_{t=1}, \pi_{t=1}) \\
&= \log\left(\frac{20}{57}\right) + \log\left(\frac{6}{19}\right) + \log\left(\frac{6}{19}\right) \\
L(X \mid \boldsymbol{p}_{t=1}, \pi_{t=1}) &\approx -3.35
\end{aligned}$$

Thankfully, we have that $L(X \mid \boldsymbol{p}_{t=1}, \pi_{t=1}) \geq L(X \mid \boldsymbol{p}_{t=0}, \pi_{t=0})$.