CSCI-567: Assignment #6

Due on Wednesday, December 2, 2015

Saket Choudhary
2170058637

Contents

Problem 1.1

Problem 1.1: (a)

$$J = \frac{1}{N}\sum_{i=1}^{N}(x_i - p_{i1}e_i - p_{i2}e_2)^T(x_i - p_{i1}e_i - p_{i2}e_2)$$

Let's define $k_i = x_i - p_{i1}$.

$$\begin{aligned}
J &= \frac{1}{N}\sum_{i=1}^{N}(k_i - p_{i2}e_2)^T(k_i - p_{i2}e_2) \\
&== \frac{1}{N}\sum_{i=1}^{N}(k_i^Tk_i - p_{i2}k_i^te_2 - p_{i2}e_2^Tk_i + p_{i2}^2e_2^Te_2) \\
\frac{\partial J}{\partial p_{i2}} &= \frac{1}{N}\sum_{i=1}^{N}(0 - k_i^Te_2 - e_2^Tk_i + 2p_{i2}e_2^Te_2) \\
&= \frac{1}{N}\sum_{i=1}^{N}(-2e_2^Tk_i + 2p_{i2})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J}{\partial p_{i2}} &= 0 \\
\implies \frac{1}{N}\sum_{i=1}^{N}(-2e_2^Tk_i + 2p_{i2}) &= 0 \\
\implies p_{i2} &= e_2^Tk_i\forall i \\
\implies p_{i2} &= e_2^T(x_i - p_{i1}e_1) \\
\implies p_{i2} &= e_2^Tx_i
\end{aligned}$$

Problem 1.1 continued on next page. . .

Problem 1.1: (b)

$$\begin{aligned}
\tilde{J} &= -e_2^T S e_2 + \lambda_2(e_2^T e_2 - 1) + \lambda_{12}(e_2^T e_1 - 0) \\
\frac{\partial \tilde{J}}{\partial e_2} &= -(S + S^T)e_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 \\
&= -2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 \text{ since } S = S^T
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \tilde{J}}{\partial e_2} &= 0 \\
\implies -2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 &= 0 \\
\implies -2e_1^T S e_2 + 2\lambda_2 e_1^T e_2 + \lambda_{12} e_1^T e_1 = 0 & \text{ pre multipllying } e_1^T \\
\implies -2(Se_1)^T e_2 + 2\lambda_2 \times 0 + \lambda_{12} \times 1 = 0 & \text{ since } S = S^T \\
\implies \lambda_{12} = 0 \text{ Since } (Se_1)^T e_2 &= 0 \\
\implies Se_2 &= \lambda_2 e_2
\end{aligned}$$

Thus, the value of e_2 minimising $\tilde{J}$ is given by second largest eigenvector of S. $Se_2 = \lambda_2 e_2$

Problem 1.2

Problem 1.2: (a)

$$\begin{aligned}
\|x_i - \sum_{j=1}^{K} p_{ij} e_j\|_2^2 &= (x_i - \sum_{j=1}^{K} p_{ij} e_j)^T (x_i - \sum_{j=1}^{K} p_{ij} e_j) \\
&= (x_i^T - \sum_{j=1}^{K} p_{ij} e_j^T)(x_i - \sum_{j=1}^{K} p_{ij} e_j) \\
&= x_i^T - x_i^T - \sum_{j=1}^{K} p_{ij} x_i^T e_j - \sum_{j=1}^{K} p_{ij} e_j^T x_i + \sum_{j,k} p_{ij} e_j^T e_k p_{ik} \\
&= x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j + \sum_{j=1}^{K} p_{ij} e_j^T e_j p_{ij} \text{ since } e_j^T e_k = 0 \text{ for } k \neq m \\
&= x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j + \sum_{j=1}^{K} e_j^T x_i x_i t^T e_j \\
&= x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j
\end{aligned}$$

Problem 1.2 continued on next page...

Problem 1.2: (a)

$$
\begin{aligned}
J_K &= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j) \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i - \sum_{j=1}^{K} e_j^T S e_j) \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i - \sum_{j=1}^{K} e_j^T \lambda_j e_j) \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i) - \frac{1}{N}\sum_{i=1}^{N}(\sum_{j=1}^{K}\lambda_j) \text{ since } e_j^T e_j = 1 \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i) - \frac{1}{N}N\sum_{j=1}^{K}\lambda_j \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^T x_i) - \sum_{j=1}^{K}\lambda_j
\end{aligned}
$$

Problem 1.2: (c)

Error from only using K components instead of D. It is easy to realise when we have $K = D$ components, $J_D = 0$ since there are no missing components. In case $K < D$, error arises due to missing components, where missing components include eigenvectors from $K + 1$ to D.

$$
\begin{aligned}
J_D &= \frac{1}{N}\sum_{i=1} x_i^T x_i - \sum_{j=1}^{D}\lambda_j \\
&= 0 \\
J_K &= \frac{1}{N}\sum_{i=1} x_i^T x_i - \sum_{j=1}^{K}\lambda_j \\
&= (\frac{1}{N}\sum_{i=1} x_i^T x_i - \sum_{j=1}^{D}\lambda_j) + \sum_{j=D+1}^{K}\lambda_j \\
&= 0 + \sum_{j=D+1}^{K}\lambda_j \\
&= \sum_{j=D+1}^{K}\lambda_j
\end{aligned}
$$

Problem 2

Problem 2: (a)

$$a = \begin{matrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{matrix}$$

Given $O = AGCGTA$
a: $P(O;\theta) = \sum_{j=1}^{2} \alpha_6(j)$
Where $\alpha_t(j) = P(O_t|S_t = j) \sum_{i=1}^{2} a_{ij}\alpha_{t-1}(j)$
$\alpha_1(j) = P(O_1|S_1 = j)P(S_1 = j)$
And for $i > 1$ $\alpha_t(j) = P(O_t|S_t = j) \times \sum_i a_{ij}\alpha_{t-1}(j)$ Thus,

$$\alpha_1(1) = P(S_1 = 1)P(O_1|S_1 = 1) = \pi_1 \times p_{1a} = 0.7 \times 0.4 = 0.28$$
$$\alpha_1(2) = \pi_2 \times p_{2a} = 0.3 \times 0.2 = 0.06$$
$$\alpha_2(1) = P(O_2|S_2 = 1) \times \sum_i a_{i1}\alpha_1(j) = b_{1g} \times (a_{11}\alpha_1(1) + a_{21}\alpha_1(2)) = 0.0992$$
$$\alpha_2(2) = b_{2g} \times (a_{12}\alpha_1(1) + a_{22}\alpha_1(2)) = 0.0184$$
$$\alpha_3(1) = b_{1c} \times (a_{11}\alpha_2(1) + a_{21}\alpha_2(2)) = 0.008672$$
$$\alpha_3(2) = b_{2c} \times (a_{12}\alpha_2(1) + a_{22}\alpha_2(2)) = 0.009264$$
$$\alpha_4(1) = b_{1g} \times (a_{11}\alpha_3(1) + a_{21}\alpha_3(2)) = 0.00425$$
$$\alpha_4(2) = b_{2g} \times (a_{12}\alpha_3(1) + a_{22}\alpha_3(2)) = 0.0014585$$
$$\alpha_5(1) = b_{1t} \times (a_{11}\alpha_4(1) + a_{21}\alpha_4(2)) = 0.000398$$
$$\alpha_5(2) = b_{2t} \times (a_{12}\alpha_4(1) + a_{22}\alpha_4(2)) = 0.0005179$$
$$\alpha_6(1) = b_{1a} \times (a_{11}\alpha_5(1) + a_{21}\alpha_5(2)) = 0.0002105$$
$$\alpha_6(2) = b_{2a} \times (a_{12}\alpha_5(1) + a_{22}\alpha_5(2)) = 0.00007810$$

$P(O;\theta) = \alpha_6(1) + \alpha_6(2) = 0.0002886$

Problem 2 continued on next page. . .

Problem 2: (b)

I refer to hidden states S_1, S_2 as 1,2 respectively. $\beta_{t-1}(i) = \sum_{j=1}^{2} \beta_t a_{ij} P(O_t|X_t = S_j)$

$$\beta_6(1) = 1$$
$$\beta_6(2) = 1$$
$$\beta_5(1) = \beta_6(1)a_{11}b_{1a} + \beta_6(2)a_{12}b_2 = 0.36$$
$$\beta_5(2) = \beta_6(1)a_{21}b_{1a} + \beta_6(2)a_{22}b_{2a} = 0.28$$
$$\beta_4(1) = \beta_5(1)a_{11}b_{1t} + \beta_5(2)a_{12}b_2 = 0.0456$$
$$\beta_4(2) = \beta_5(1)a_{21}b_{1t} + \beta_5(2)a_{22}b_{2a} = 0.0648$$

$$P(X_6 = S_i|O,\theta) = \frac{\alpha_6(S_i)\beta_6(S_i)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)}$$

Thus,

$$\begin{aligned} P(X_6 = S_1|O,\theta) &= \frac{\alpha_6(S_1)\beta_6(S_1)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} \\ &= \frac{0.0002105}{0.0002886} \\ &= 0.7293 \\ P(X_6 = S_2|O,\theta) &= \frac{\alpha_6(S_2)\beta_6(S_2)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} \\ &= \frac{0.000078}{0.0002886} \\ &= 0.2702 \end{aligned}$$

Problem 2: (b)

$$\begin{aligned} P(X_4 = S_1|O,\theta) &= \frac{\alpha_4(S_1)\beta_4(S_1)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} \\ &= \frac{0.0456}{0.1104} \\ &= 0.413 \\ P(X_4 = S_2|O,\theta) &= \frac{\alpha_4(S_2)\beta_4(S_2)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} \\ &= \frac{0.0648}{0.1104} \\ &= 0.587 \end{aligned}$$

Problem 2 continued on next page. . .

Problem 2: (d)

$\delta_t(j) = \max_i \delta_{t-1}(i)a_{ij}P(x_t|Z_t = s_i)$

$$\delta_1(1) = \pi_1 b_{1a} = 0.28$$
$$\delta_1(2) = \pi_2 b_{2a} = 0.06$$
$$\delta_2(1) = b_{1g} \times max(\delta_1(1)a_{11}, \delta_1(2)a_{21}) = 0.896$$
$$\delta_2(2) = b_{2g} \times max(\delta_1(1)a_{12}, \delta_1(2)a_{22}) = 0.072$$
$$\delta_3(1) = b_{1c} \times (\delta_2(1)a_{11} + \delta_2(2)a_{21})$$

Problem 2: (e)

$O_7 = \arg\max_O P(O|O\theta)$

$$\begin{aligned}
P(O_7|O) &= \sum_{i=1}^{2} P(O_7, X_7 = S_i|O) \\
&= \sum_{i=1}^{2} P(O_7|X_7 = S_i) \times \sum_{j=1}^{2} P(X_7 = S_i, X_6 = S_j|O) \\
&= \sum_{i=1}^{2} P(O_7|X_7 = S_i) \times \sum_{j=1}^{2} P(X_7 = S_i|X_6 = S_j)P(X_6 = S_j|O) \\
&= b_{1k} \times (P(X_6 = S_1|\theta) \times a_{11} + P(X_6 = S_2|\theta) \times a_{21}) \\
&+ b_{2k} \times (P(X_6 = S_1|\theta) \times a_{12} + P(X_6 = S_2|\theta) \times a_{22}) \; where \; k \in (A, C, T, G)
\end{aligned}$$

Define $c_1 = P(X_6 = S_1|\theta) \times a_{11} + P(X_6 = S_2|\theta) \times a_{21} = 0.6915$
Define,$c_2 = P(X_6 = S_1|\theta) \times a_{12} + P(X_6 = S_2|\theta) \times a_{22} = 0.307$
Thus, $P(O_7 = k|\theta) = 0.6915b_{1k} + 0.307b_{2k}$

$$P(O_7 = A|\theta) = 0.6915 * 0.4 + 0.307 * 0.2 = 0.338$$
$$P(O_7 = T|\theta) = 0.6915 * 0.1 + 0.307 * 0.3 = 0.16125$$
$$P(O_7 = C|\theta) = 0.6915 * 0.1 + 0.307 * 0.3 = 0.16125$$
$$P(O_7 = G|\theta) = 0.6915 * 0.4 + 0.307 * 0.2 = 0.338$$

Thus, A, G are equiprobable as 7^{th} observed sequence.

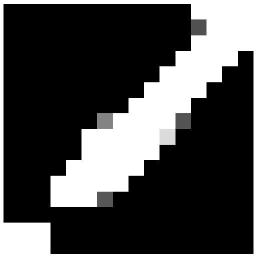
Figure 1: PC-1

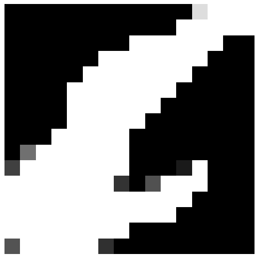
Figure 2: PC-2

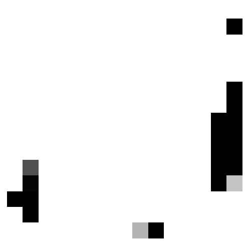
Figure 3: PC-3

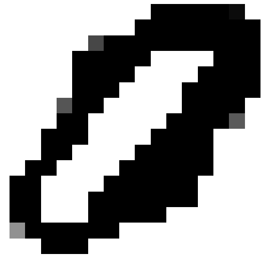
Figure 4: PC-4

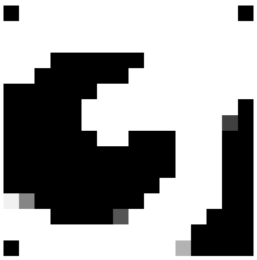
Figure 5: PC-5


Figure 6: PC-6

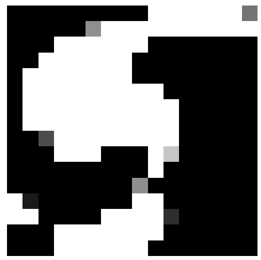
Figure 7: PC-7

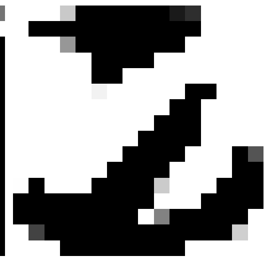
Figure 8: PC-8

Problem 3.1

Problem 3.1: (b)

Problem 3.1 continued on next page...

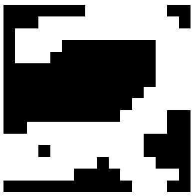
Figure 9: #5500 Original

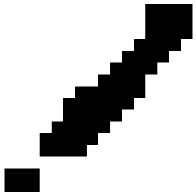
Figure 10: #5500 K-1

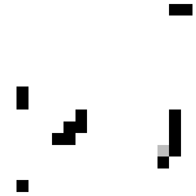
Figure 11: #5500 K-5

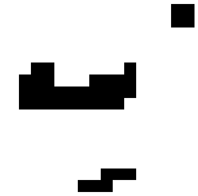
Figure 12: #5500 K-10

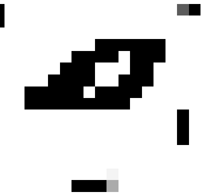
Figure 13: #5500 K-20


Figure 14: #5500 K-80

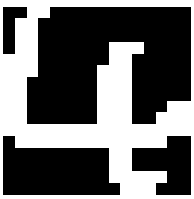
Figure 15: #6500 Original

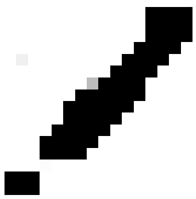
Figure 16: #6500 K-1

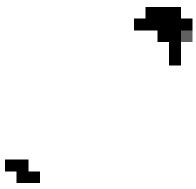
Figure 17: #6500 K-5

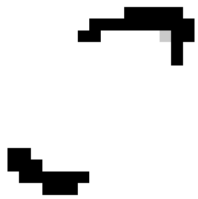
Figure 18: #6500 K-10

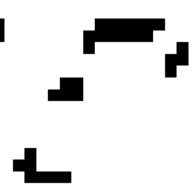
Figure 19: #6500 K-20

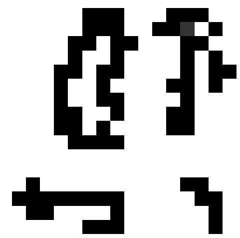
Figure 20: #6500 K-80

Problem 3.1: (c)

Problem 3.1: (d)

#PC	Training Accuracy	Test Accuracy	Time taken(s)
1	0.516778	0.122500	32.125276
5	0.890889	0.434000	17.126630
10	0.947333	0.635500	14.667870
20	0.947556	0.742500	14.161249
80	0.920444	0.763500	17.982115

Thus we see that the training accuracy increases as the number pf principal components increase. This is intuitive since we have more information with higher principal components. Training accuracy decreases for 80 principal components over 20 indicating that the next 60 principal components are not distinctive enough. The testing accuracy continuously increases as number of principal components increase indicating higher information gain by increasing the number of principal components.

The time taken by Decision tree classifier decreases as the number of principal components increase. This is expected since there is limited information available with say 1 principal component and hence the decision tree will keep on growing in depth. whereas when higher number of PCs are used the tree depth will be smaller since there is more information at each branching step.

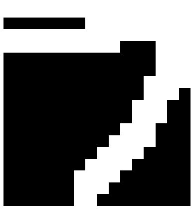
Figure 21: #7500 Original

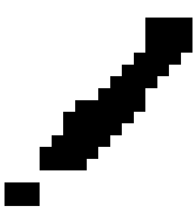
Figure 22: #7500 K-1

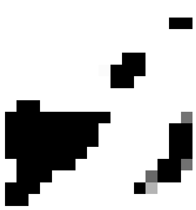
Figure 23: #7500 K-5

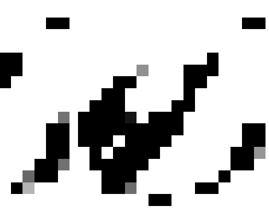
Figure 24: #7500 K-10

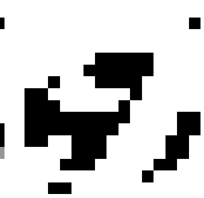
Figure 25: #7500 K-20

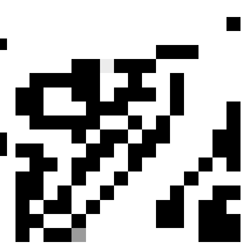
Figure 26: #7500 K-80

Figure 27: #8000 Original

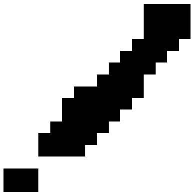

Figure 28: #8000 K-1

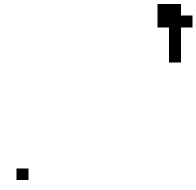

Figure 29: #8000 K-5

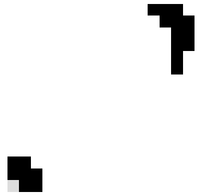

Figure 30: #8000 K-10

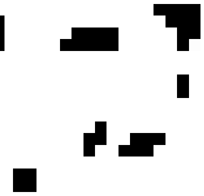

Figure 31: #8000 K-20

Figure 32: #8000 K-80

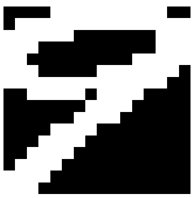

Figure 33: #8500 Original

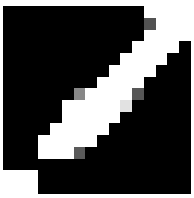

Figure 34: #8500 K-1

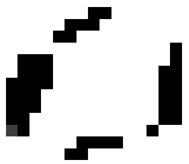

Figure 35: #8500 K-5

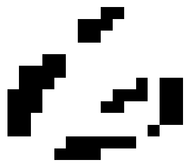

Figure 36: #8500 K-10

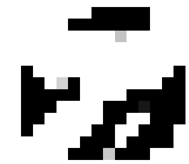

Figure 37: #8500 K-20

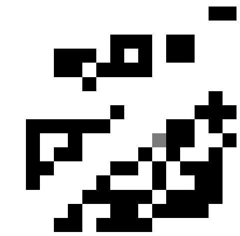

Figure 38: #8500 K-80

Problem 3.2

Problem 3.2: (a)

Length of shortest trace: 109
Length of longest trace: 435
How many different observations: 22

Problem 3.2: (a)

Hidden States: S
Observations: $O^{(1)}, O^{(2)} \dots O^{(D)}$ where $O^{(i)}$ denotes the the i^{th} training sample. $\theta = \{\pi_i, a_{ij}, b_{ik}\}$
Let T_i denote the trace length of i^{th} training sample.
Thus, any observation $O^{(i)} = \{O_1^{(i)}, O_2^{(i)} \dots O_{T_i}^{(i)}\}$

$$
\begin{aligned}
L_D(\theta, A, B) &= \sum_{i=1}^{D} \log(P(O_1^{(i)}, O_1^{(i)}, \dots O_{T_i}^{(i)}, S_1^{(i)}, S_2^{(i)}, \dots S_{T_i}^{(i)})) \\
&= \sum_{i=1}^{D} (\log(P(S_1^{(i)}, S_2^{(i)}, \dots S_{T_i}^{(i)}) + P(O_1^{(i)}, O_1^{(i)}, \dots O_{T_i}^{(i)} | S_1^{(i)}, S_2^{(i)}, \dots S_{T_i}^{(i)}))) \\
&= \sum_{i=1}^{D} \log(P(S_1^{(i)}) + \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(P(S_t^{(i)} | S_{t-1}^{(i)})) + \sum_{i=1}^{D} \sum_{t=1}^{T_i} \log(P(O_t^{(i)} | S_t^{(i)}))) \\
&= \sum_{i=1}^{D} \log(P(S_1^{(i)}) + \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(P(S_t^{(i)} | S_{t-1}^{(i)})) + \sum_{i=1}^{D} \sum_{t=1}^{T_i} \log(P(O_t^{(i)} | S_t^{(i)}))) \\
&= \sum_{i=1}^{D} \log(\pi_{S_1}^{(i)}) + \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(A_{t-1,t}^{(i)} + \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(B_{S_t^{(i)} O_t^{(i)}}^{(i)})
\end{aligned}
$$

Note $B_{S_t^{(i)} O_t^{(i)}}^{(i)}$ denotes the emission probability when state is $S_t^{(i)}$ and emission is $O_t^{(i)}$

Problem 3.2: (a)

Let S' denote the set of all possiblities that $S_t^{(i)}$ can take,

$$Q(\theta, \theta^s) = \sum_{S \in S'} L_D(\theta, s, A, B) \times P(s|O;\theta)$$
$$= \sum_{S \in S'} \sum_{i=1}^{D} \log(\pi_{S_1}^{(i)}) + \sum_{S \in S'} \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(A_{t-1,t}^{(i)} + \sum_{S \in S'} \sum_{i=1}^{D} \sum_{t=2}^{T_i} \log(B^{(i)}_{S_t^{(i)} O_t^{(i)}})$$

Now define the total number of hidden states to be M.
Given constraints (for each training dataset):

$$\sum_i \pi_i = 1$$
$$\sum_j A_{ij} = 1$$
$$\sum_j B_{ij} = 1$$

Then we define a langragian as follows:

$$L(\theta, \theta^s) = Q(\theta, \theta^s) - \lambda_\pi (\sum_{i=1}^{M} \pi_i - 1) + \sum)i = 1^M \lambda_{A_i} (\sum_{j=1}^{M} \lambda A_{ij} - 1) - \sum_{i=1}^{M} \lambda_{B_i} (\sum_{j=1}^{M} B_{ij} - 1)$$

Now,

$$\frac{\partial L}{\partial \pi_{S_i}} = 0$$
$$= \frac{\partial}{\partial \pi_{S_i}} (\sum_{S \in S'} \sum_{i=d}^{D} \sum_{j=1}^{M} \log(\pi_{S_i}) P(S_1^{(i)} = j, O; \theta) - \lambda_\pi = 0$$
$$= \sum_{i=1}^{D} \frac{P(S_1^{(i)} = j, O; \theta)}{\pi_{S_i}} - \lambda_\pi = 0$$

Since the likelihood is linear in terms of probability, the M step involves a posterior calculation:

$$\pi_{S_i} = \frac{\sum_{i=1}^{D} P(S_1^{(i)} = S_i, O; \theta)}{\sum_{i=d}^{D} \sum_{j=1}^{M} \sum_{i=1}^{D} P(S_1^{(i)} = S_i, O; \theta)}$$