Problem 1

1.1

(a)

Given cost function:

$$J = \frac{1}{N}\sum_{i=1}^{N}(x_i - p_{i1}e_1 - p_{i2}e_2)^T(x_i - p_{i1}e_1 - p_{i2}e_2) \tag{1}$$

$$||e_1|| = e_1^T e_1 = 1 \tag{2}$$

$$||e_2|| = e_2^T e_2 = 1 \tag{3}$$

$$e_1^T e_2 = e_2^T e_1 = 0 \tag{4}$$

For any instance i, using above equations:

$$\begin{aligned}\frac{\partial J}{\partial p_{i2}} &= \frac{\partial((x_i - p_{i1}e_1 - p_{i2}e_2)^T(x_i - p_{i1}e_1 - p_{i2}e_2))}{\partial p_{i2}} \\ &= 2e_{i2}^T(x_i - p_{i1}e_1 - p_{i2}e_2) \\ &= 2(e_{i2}^T x_i - p_{i1}e_{i2}^T e_1 - p_{i2}e_{i2}^T e_2) \\ &= 2(e_{i2}^T x_i - p_{i2}) = 0\end{aligned}$$

$$\Rightarrow p_{i2} = e_2^T x_i$$

(b)

Given cost function:

$$\tilde{J} = -e_2^T S e_2 + \lambda_2(e_2^T e_2 - 1) + \lambda_{12}(e_2^T e_1 - 0) \tag{5}$$

Lets find the value of e_2 that minimizes the above cost function:

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

$$\Rightarrow \hat{e_2} = \tfrac{1}{2}(S - I\lambda)^{-1}\lambda_{12}e_1$$

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

$$\Rightarrow e_2^T e_2 = 1$$

$$\frac{\partial \tilde{J}}{\partial \lambda_{12}} = \frac{\partial(-e_2^T S e_2 + \lambda_2(e_2^T e_2 - 1) + \lambda_{12}(e_2^T e_1 - 0))}{\partial \lambda_{12}}$$
$$= e_2^T e_1 - 0 = 0$$
$$\Rightarrow e_2^T e_1 = 0$$

1.2

(a)

Given,

$$e_j^T e_j = 1 \quad (6)$$
$$e_j^T e_m = 0 \quad \forall j \neq m \quad (7)$$
$$p_{ij} = e_j^T x_i \quad (8)$$

$$||x_i - \sum_{j=1}^{K} p_{ij} e_j||_2^2 = ||x_i - \sum_{j=1}^{K} e_j^T x_i e_j||_2^2$$
$$= (x_i - \sum_{j=1}^{K} e_j^T x_i e_j)^T (x_i - \sum_{j=1}^{K} e_j^T x_i e_j)$$
$$= [x_i^T - \sum_{j=1}^{K} e_j^T x_i^T e_j][x_i - \sum_{j=1}^{K} e_j^T x_i e_j]$$
$$= [x_i^T - (e_1^T x_i^T e_1 + \ldots + e_K^T x_i^T e_K)][x_i - (e_1^T x_i e_1 + \ldots + e_K^T x_i e_K)]$$
$$= x_i^T x_i - 2(e_1^T x_i x_i^T e_1 + \ldots + e_K^T x_i x_i^T e_K) + (e_1^T x_i^T e_1 + \ldots + e_K^T x_i^T e_K) * (e_1^T x_i e_1 + \ldots + e_K^T x_i e_K)$$
$$= x_i^T x_i - 2\sum_{j=1}^{K} e_j^T x_i x_i^T e_j + (e_1^T x_i^T e_1 + \ldots + e_K^T x_i^T e_K) * (e_1^T x_i e_1 + \ldots + e_K^T x_i e_K)$$
$$= x_i^T x_i - 2\sum_{j=1}^{K} e_j^T x_i x_i^T e_j + (\sum_{j=1}^{K}\sum_{m=1}^{K} e_j^T x_i e_j^T e_m x_i^T e_j)$$

{Using properties (6) and (7), most terms zero out}

$$x_i^T x_i - 2\sum_{j=1}^{K} e_j^T x_i x_i^T e_j + \sum_{j=1}^{K} e_j^T x_i x_i^T e_j$$
$$\Rightarrow ||x_i - \textstyle\sum_{j=1}^{K} p_{ij} e_j||_2^2 = x_i^T x_i - \textstyle\sum_{j=1}^{K} e_j^T x_i x_i^T e_j$$

(b)

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

(c)

When $K = D$, $J_D = 0$. So we have:

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

Problem 2

(a)

$$\mathbf{P}(\mathbf{O};\boldsymbol{\Theta})$$

$$\alpha_1(1) = b_{1A}\pi_1 = 0.4 \times 0.7 = 0.28$$

$$\alpha_1(2) = b_{2A}\pi_2 = 0.2 \times 0.3 = 0.06$$

$$\alpha_2(1) = b_{1G}(\alpha_1(1)a_{11} + \alpha_1(2)a_{21}) = 0.4 \times (0.28 \times 0.8 + 0.06 \times 0.4) = 0.0992$$

$$\alpha_2(2) = b_{2G}(\alpha_1(1)a_{12} + \alpha_1(2)a_{22}) = 0.2 \times (0.28 \times 0.2 + 0.06 \times 0.6) = 0.0184$$

$$\alpha_3(1) = b_{1C}(\alpha_2(1)a_{11} + \alpha_2(2)a_{21}) = 0.1 \times (0.0992 \times 0.8 + 0.0184 \times 0.4) = 0.008672$$

$$\alpha_3(2) = b_{2C}(\alpha_2(1)a_{12} + \alpha_2(2)a_{22}) = 0.3 \times (0.0992 \times 0.2 + 0.0184 \times 0.6) = 0.009264$$

$$\alpha_4(1) = b_{1G}(\alpha_3(1)a_{11} + \alpha_3(2)a_{21}) = 0.4 \times (0.008672 \times 0.8 + 0.009264 \times 0.4) = 0.00425728$$

$$\alpha_4(2) = b_{2G}(\alpha_3(1)a_{12} + \alpha_3(2)a_{22}) = 0.2 \times (0.008672 \times 0.2 + 0.009264 \times 0.6) = 0.00145856$$

$$\alpha_5(1) = b_{1T}(\alpha_4(1)a_{11} + \alpha_4(2)a_{21}) = 0.1 \times (0.00425728 \times 0.8 + 0.00145856 \times 0.4) = 0.0003989248$$

$$\alpha_5(2) = b_{2T}(\alpha_4(1)a_{12} + \alpha_4(2)a_{22}) = 03 \times (0.00425728 \times 0.2 + 0.00145856 \times 0.6) = 0.0005179776$$

$$\alpha_6(1) = b_{1A}(\alpha_5(1)a_{11} + \alpha_5(2)a_{21}) = 0.4 \times (0.0003989248 \times 0.8 + 0.0005179776 \times 0.4) = 0.000210532352$$

$$\alpha_6(2) = b_{2A}(\alpha_5(1)a_{12} + \alpha_5(2)a_{22}) = 0.2 \times (0.0003989248 \times 0.2 + 0.0005179776 \times 0.6) = 0.00007811$$

$$P(O;\Theta) = \alpha_6(1) + \alpha_6(2) = 0.000210532352 + 0.00007811 = 0.000288642352$$

(b)

$$\mathbf{P(X_6 = S_i|O;\Theta)}$$
$$\beta_6(1) = 1$$
$$\beta_6(2) = 1$$
$$\beta_5(1) = (b_{1A}a_{11}\beta_6(1) + b_{2A}a_{12}\beta_6(2)) = (0.4 \times 0.8 \times 1 + 0.2 \times 0.2 \times 1) = 0.36$$
$$\beta_5(2) = (b_{1A}a_{21}\beta_6(1) + b_{2A}a_{22}\beta_6(2)) = (0.4 \times 0.4 \times 1 + 0.2 \times 0.6 \times 1) = 0.28$$
$$\beta_4(1) = (b_{1T}a_{11}\beta_5(1) + b_{2T}a_{12}\beta_5(2)) = (0.1 \times 0.8 \times 0.36 + 0.3 \times 0.2 \times 0.28) = 0.0456$$
$$\beta_4(2) = (b_{1T}a_{21}\beta_5(1) + b_{2T}a_{22}\beta_5(2)) = (0.1 \times 0.4 \times 0.36 + 0.3 \times 0.6 \times 0.28) = 0.0648$$
$$\beta_3(1) = (b_{1G}a_{11}\beta_4(1) + b_{2G}a_{12}\beta_4(2) = (0.4 \times 0.8 \times 0.0456 + 0.2 \times 0.2 \times 0.0648) = 0.017184$$
$$\beta_3(2) = (b_{1G}a_{21}\beta_4(1) + b_{2G}a_{22}\beta_4(2) = (0.4 \times 0.4 \times 0.0456 + 0.2 \times 0.6 \times 0.0648) = 0.015072$$
$$\beta_2(1) = (b_{1C}a_{11}\beta_3(1) + b_{2C}a_{12}\beta_3(2)) = (0.1 \times 0.8 \times 0.017184 + 0.3 \times 0.2 \times 0.015072) = 0.00227904$$
$$\beta_2(2) = (b_{1C}a_{21}\beta_3(1) + b_{2C}a_{22}\beta_3(2)) = (0.1 \times 0.4 \times 0.017184 + 0.3 \times 0.6 \times 0.015072) = 0.00340032$$
$$\beta_1(1) = (b_{1G}a_{11}\beta_2(1) + b_{2G}a_{12}\beta_2(2)) = (0.4 \times 0.8 \times 0.00227904 + 0.2 \times 0.2 \times 0.00340032) = 0.0008653056$$
$$\beta_1(2) = (b_{1G}a_{21}\beta_2(1) + b_{2G}a_{22}\beta_2(2)) = (0.4 \times 0.4 \times 0.00227904 + 0.2 \times 0.6 \times 0.00340032) = 0.0007726848$$
$$P(X_6 = S_1|O;\Theta) = \frac{\alpha_6(1)\beta_6(1)}{\alpha_6(1)\beta_6(1) + \alpha_6(2)\beta_6(2)} = \frac{0.000210532352 \times 1}{0.000210532352 \times 1 + 0.00007811 \times 1}$$
$$= \mathbf{0.72938}$$
$$P(X_6 = S_2|O;\Theta) = 1 - P(X_6 = S_1|O;\Theta) = 1 - 0.72938 = \mathbf{0.27062}$$

(c)

$$\mathbf{P(X_4 = S_i|O;\Theta)}$$
$$P(X_4 = S_1|O;\Theta) = \frac{\alpha_4(1)\beta_4(1)}{\alpha_4(1)\beta_4(1) + \alpha_4(2)\beta_4(2)}$$
$$= \frac{0.00425728 \times 0.0456}{0.00425728 \times 0.0456 + 0.00145856 \times 0.0648}$$
$$= \mathbf{0.672559}$$
$$P(X_4 = S_2|O;\Theta) = 1 - P(X_4 = S_1|O;\Theta) = 1 - 0.78790 = \mathbf{0.327441}$$

(d)

$$P(X_1 = S_1|O;\Theta) = \frac{\alpha_1(1)\beta_1(1)}{\alpha_1(1)\beta_1(1) + \alpha_1(2)\beta_1(2)} = \frac{0.28 \times 0.0008653056}{0.28 \times 0.0008653056 + 0.06 \times 0.0007726848} = \mathbf{0.83938}$$

$$P(X_1 = S_2|O;\Theta) = \frac{\alpha_1(2)\beta_1(2)}{\alpha_1(1)\beta_1(1) + \alpha_1(2)\beta_1(2)} = 1 - P(X_1 = S_1|O;\Theta) = \mathbf{0.16062}$$

$$P(X_2 = S_1|O;\Theta) = \frac{\alpha_2(1)\beta_2(1)}{\alpha_2(1)\beta_2(1) + \alpha_2(2)\beta_2(2)} = \frac{0.0992 \times 0.00227904}{0.0992 \times 0.00227904 + 0.0184 \times 0.00340032} = \mathbf{0.78324}$$

$$P(X_2 = S_2|O;\Theta) = \frac{\alpha_2(2)\beta_2(2)}{\alpha_2(1)\beta_2(1) + \alpha_2(2)\beta_2(2)} = 1 - P(X_2 = S_1|O;\Theta) = \mathbf{0.21676}$$

$$P(X_3 = S_1|O;\Theta) = \frac{\alpha_3(1)\beta_3(1)}{\alpha_3(1)\beta_3(1) + \alpha_3(2)\beta_3(2)} = \frac{0.008672 \times 0.017184}{0.008672 \times 0.017184 + 0.009264 \times 0.015072} = \mathbf{0.51627}$$

$$P(X_3 = S_2|O;\Theta) = \frac{\alpha_3(2)\beta_3(2)}{\alpha_3(1)\beta_3(1) + \alpha_3(2)\beta_3(2)} = 1 - P(X_3 = S_1|O;\Theta) = \mathbf{0.48373}$$

$$P(X_5 = S_1|O;\Theta) = \frac{\alpha_5(1)\beta_5(1)}{\alpha_5(1)\beta_5(1) + \alpha_5(2)\beta_5(2)} = \frac{0.0003989248 \times 0.36}{0.0003989248 \times 0.36 + 0.0005179776 \times 0.28} = \mathbf{0.49753}$$

$$P(X_5 = S_2|O;\Theta) = \frac{\alpha_5(2)\beta_5(2)}{\alpha_5(1)\beta_5(1) + \alpha_5(2)\beta_5(2)} = 1 - P(X_5 = S_1|O;\Theta) = \mathbf{0.50247}$$

Hence, the Most likely sequence is $S_1S_1S_1S_1S_2S_1$

(e)

$$\mathbf{P(O_7|O;\Theta)}$$

A:

$$\begin{aligned} P(O_7 = A|O;\Theta) &= P(X_6 = S_1|O;\Theta) \times b_{1A} + P(X_6 = S_2|O;\Theta) \times b_{2A} \\ &= 0.72938 \times 0.4 + 0.27062 \times 0.2 = \mathbf{0.345876} \end{aligned}$$

C:

$$\begin{aligned} P(O_7 = C|O;\Theta) &= P(X_6 = S_1|O;\Theta) \times b_{1C} + P(X_6 = S_2|O;\Theta) \times b_{2C} \\ &= 0.72938 \times 0.1 + 0.27062 \times 0.3 = \mathbf{0.154124} \end{aligned}$$

G:

$$\begin{aligned} P(O_7 = G|O;\Theta) &= P(X_6 = S_1|O;\Theta) \times b_{1G} + P(X_6 = S_2|O;\Theta) \times b_{2G} \\ &= 0.72938 \times 0.4 + 0.27062 \times 0.2 = \mathbf{0.345876} \end{aligned}$$

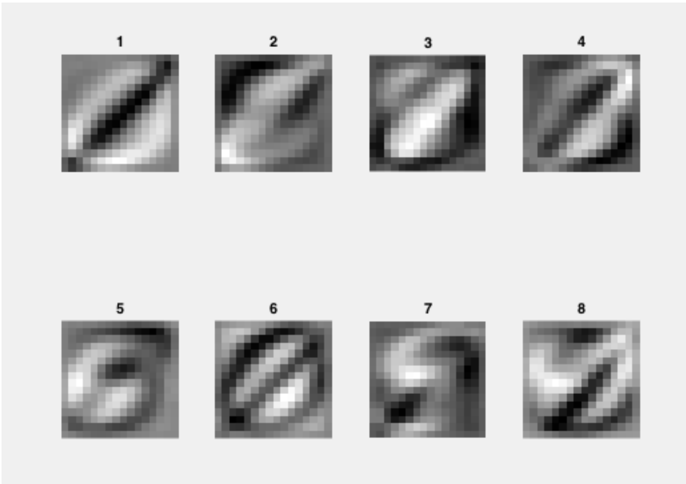

Figure 1: 3.1(b) Eigendigits of top 8 eigenvalues

T:

$$\begin{aligned} P(O_7 = T|O;\Theta) &= P(X_6 = S_1|O;\Theta) \times b_{1T} + P(X_6 = S_2|O;\Theta) \times b_{2T} \\ &= 0.72938 \times 0.1 + 0.27062 \times 0.3 = \mathbf{0.154124} \end{aligned}$$

Hence, we say A and G have equal probability for O_7

Problem 3

3.1

(a)

Implemented *get_sorted_eigenvecs.m* that returns D eigenvectors.

(b)

(c)

(d)

K	**Train Accuracy**	**Test Accuracy**	**Runtime**
1	51.6778	23.9000	44.9400
5	89.0889	65.5000	24.7700
10	94.7333	79.6500	20.5600
20	94.7556	81.5500	20.1600
80	92.0444	78.9500	24.3500

Analysis

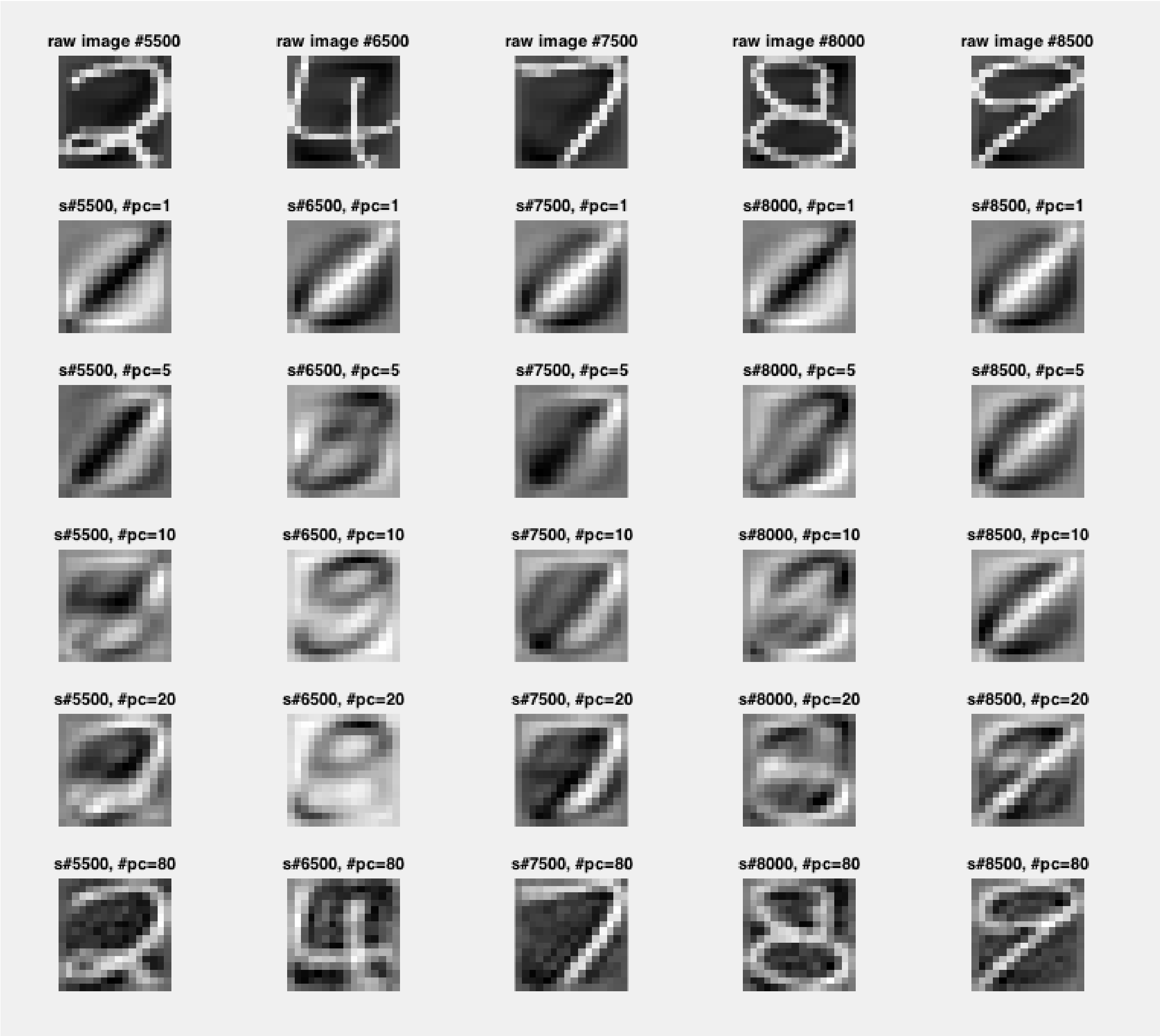

Figure 2: 3.1(c) Eigenfaces for various input data and K

For both Train and Test accuracy, we observe that the accuracy increases as the value of K increases and after reaching a plateau, the accuracy starts to reduce. This is because, higher the value of K, higher is the information retained in the reconstructed data. At the same time, once it hits the plateau, model starts to overfit thus reducing the accuracy.

For Runtime, we see that it reduces as the value of K increases and after reaching a plateau the runtime starts to increase again. This is because, the decision trees starts to become simpler as K increases. At the same time, once it hits the plateau, decision trees start growing complex again and hence the runtime starts to increase too.

3.2

3.2.1

The shortest trace is **109** for **PID: 738**
The longest trace is **435** for **PID: 4121 and 4190**
We totally found **15 unique PIDs**

3.2.2

Lets assume multiple sequences is given by $X = \{x_1, \ldots x_n\}$ which correspond to Hidden states $Z = \{z_1, \ldots, z_n\}$. Each sequence is of length T. Then the Likelihood is given by :

$$L(\theta, \gamma) = \prod_{x1,x2,\ldots,x_n} \prod_{t=1}^{T} p(x_{nt})$$

$$= \prod_{x1,x2,\ldots,x_n} p(x_{n0}) \prod_{t=1}^{T} p(x_{nt}|x_{n(t-1)}; \theta) P(z_t|x_{nt}; \gamma)$$

Applying log on both sides:

$$l(\theta, \gamma) = log \left(\sum_{x1,x2,\ldots,x_n} p(x_{n0}) \prod_{t=1}^{T} p(x_{nt}|x_{n(t-1)}; \theta) P(z_t|x_{nt}; \gamma) \right)$$

$$l(\theta, \gamma) = log \left(\sum_{x1,x2,\ldots,x_n} p(x_{n0}) \prod_{t=1}^{T} \theta_{x_{nt}|x_{n(t-1)}} \prod_{t=1}^{T} \gamma_{z_t|x_{nt}} \right)$$

3.2.3

E Step:

$$l(\theta, \gamma) = max_{\theta,\gamma} \sum_{n=1}^{N} \sum_{t=0}^{T} \sum_{x_{nt}, xn(t+1)} q(x_{nt}, x_{n(t+1)}) logp(x_{n(t+1)}|x_{nt}; \theta) + \sum_{n=1}^{N} \sum_{t=0}^{T} \sum_{x_{nt}} q(x_{nt}) logp(z_t|x_{nt}; \gamma)$$

θ and γ computed from Soft counts:

$$n_{(i,j)} = \sum_{t=0}^{T-1} q(x_{t+1} = i, x_t = j)$$
$$m_{(k,l)} = \sum_{t=0}^{T} q(z_t = ki, x_t = l)$$

By differentiating, we get:

$$\theta_{i|j} = \frac{n_{(i,j)}}{\sum_{i=1} n_{(i,j)}}$$
$$\gamma_{k|l} = \frac{m_{(k,l)}}{\sum_{k=1}^{K} m_{(k,l)}}$$