CSCI567 Fall 2015

Homework #6

Vaibhav Behl

3123054410

vbehl@usc.edu

1. Principal Component Analysis

1.1 Derivation of Second Principal Component

a) Given cost function:

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

This can be written as:

$$= \frac{1}{N}\sum \left|\left|x_i - p_{i1}e_1 - p_{i2}e_2\right|\right|^2$$

Taking partial derivative:

$$\frac{\delta J}{\delta p_{i2}} = -\frac{1}{N}\sum 2\, e_2^T.(x_i - p_{i1}e_1 - p_{i2}e_2) = 0$$

Simplifying and using $\left|\left|e_2\right|\right|_2 = 1 \; and \; e_1^T e_2 = 0$:

$$= \frac{2}{N}\sum(e_2^T x_i - 0 - p_{i2}) = 0$$

Thus we get:

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

b) Given cost function:

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

Taking derivative w.r.t. e_2:

$$\frac{\delta J}{\delta e_2} = -Se_2 + \lambda_2 e_2 + \lambda_{12} e_1 = 0$$

Multiplying e_1 to the left side of this equation:

$$-\,e_1^T S e_2 + \lambda_2 e_1^T e_2 + \lambda_{12} e_1^T e_1 = 0$$
$$-\,0 + 0 + \lambda_{12} = 0$$

This means that:

$$Se_2 - \lambda_2 e_2 = 0$$

This is also an eigenvalue equation, and as it can be seen, the value of e_2 which minimizes this is given by an eigenvector associated with the second largest eigenvalue of S.

1.2 Derivation of PCA Residual Error

a) We want to prove:

$$\left|\left|x_i - \sum_{j=1}^{K} p_{ij}e_j\right|\right|_2^2 = x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j$$

Using induction we first prove for $K = 1$:
The LHS is as follows:

$$\left||x_i - p_{i1}e_1\right||^2$$
$$= x_i^T x_i + p_{i1}^2 e_1^T e_1 - 2p_{i1}x_i^T e_1$$
$$= x_i^T x_i + p_{i1}^2 - 2p_{i1}^2$$
$$= x_i^T x_i - p_{i1}^2$$

The RHS is:

$$x_i^T x_i - e_1^T x_i x_i^T e_1$$
$$= x_i^T x_i - e_1^T x_i (e_1^T x_i)^T$$
$$= x_i^T x_i - p_{i1}^2$$

Hence proved for $K = 1$.
Now assuming this is true for $K = k - 1$, Taking LHS,

$$\left\|x_i - \sum_{j=1}^{K} p_{ij}e_j\right\|_2^2$$

Splitting summation,

$$= \left\|x_i - \sum_{j=1}^{K-1} p_{ij}e_j + p_{ik}e_k\right\|_2^2$$

$$= \left\|x_i - \sum_{j=1}^{K-1} p_{ij}e_j\right\|_2^2 + (p_{ik}e_k)^2 - 2p_{ik}e_k(x_i - \sum_{j=1}^{K-1} p_{ij}e_j)$$

Substituting value for $K = k - 1$,

$$= x_i^T x_i - \sum_{j=1}^{k-1} e_j^T x_i x_i^T e_j + e_k^T x_i x_i^T e_k - 2p_{ik}e_k x_i + 2p_{ik}e_k \sum_{j=1}^{K-1} p_{ij}e_j$$

Last two terms are zero, since vectors are orthogonal.
Hence,

$$\left\|x_i - \sum_{j=1}^{K} p_{ij}e_j\right\|_2^2 = x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j$$

b) Starting with LHS:

$$J_K = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^T x_i - \sum_{j=1}^{K} e_j^T x_i x_i^T e_j\right)$$

Taking sum inside,

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - \sum_{i=1}^{N}\sum_{j=1}^{K} e_j^T x_i x_i^T e_j$$

Interchanging the summations, using the property mentioned, and solving:

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - \sum_{j=1}^{K} e_j^T S e_j$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - \sum_{j=1}^{K} \lambda_j$$

c) For $K = D, J_D = 0$,

$$J_D = \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - \sum_{j=1}^{D} \lambda_j = 0$$

$$\sum_{j=1}^{D} \lambda_j = \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i$$

Now, J_K can be written as:

$$J_K = \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - \sum_{j=1}^{K} \lambda_j$$

$$= \sum_{j=1}^{D} \lambda_j - \sum_{j=1}^{K} \lambda_j$$

Hence,

$$J_K = \sum_{j=K+1}^{D} \lambda_j$$

2. Hidden Markov Model

Given,

$$\pi_1 = 0.7, \pi_2 = 0.3$$
$$a_{11} = 0.8, a_{12} = 0.2, a_{21} = 0.4, a_{22} = 0.6$$
$$b_{1A} = 0.4\ , b_{1C} = 0.1, b_{1G} = 0.4, b_{1T} = 0.1$$
$$b_{2A} = 0.2, b_{2C} = 0.3, b_{2G} = 0.2, b_{2T} = 0.3$$

Sequence,

$$O = AGCGTA$$

Hidden states- S_1, S_2

a) Probability of $P(O_{1:6}; \theta)$. We'll calculate this using Forward algorithm.
   Base Case:

$$\alpha_1(j) = \pi_j b_{jO_1}$$
$$= \pi_j b_{jA}$$

   Recursion:

$$\alpha_t(j) = P(x_t | Z_t = s_j) \sum_i a_{ij} \alpha_{t-1}(i)$$
$$= b_{jO_t} \sum_i a_{ij} \alpha_{t-1}(i)$$

   Now for $(j = 1,2)$:

$$\alpha_1(1) = 0.7 * 0.4 = 0.28$$
$$\alpha_1(2) = 0.3 * 0.2 = 0.06$$

$$\alpha_2(1) = 0.4 * (0.8 * 0.28 + 0.4 * 0.06) = 0.0992$$
$$\alpha_2(2) = 0.2 * (0.2 * 0.28 + 0.6 * 0.06) = 0.0184$$

$$\alpha_3(1) = 0.1 * (0.8 * 0.0992 + 0.4 * 0.0184) = 0.0087$$
$$\alpha_3(2) = 0.3 * (0.2 * 0.0992 + 0.6 * 0.0184) = 0.0093$$

$$\alpha_4(1) = 0.4 * (0.8 * 0.0087 + 0.4 * 0.0093) = 0.0043$$
$$\alpha_4(2) = 0.2 * (0.2 * 0.0087 + 0.6 * 0.0093) = 0.0015$$

$$\alpha_5(1) = 0.1 * (0.8 * 0.0043 + 0.4 * 0.0015) = 0.0004$$
$$\alpha_5(2) = 0.3 * (0.2 * 0.0043 + 0.6 * 0.0015) = 0.0005$$

$$\alpha_6(1) = 0.4 * (0.8 * 0.0004 + 0.4 * 0.0005) = 0.0002$$
$$\alpha_6(2) = 0.2 * (0.2 * 0.0004 + 0.6 * 0.0005) = 0.000076$$

   Now,

$$P(O_{1:6}; \theta) = \alpha_6(1) + \alpha_6(2) = 0.000276$$

b) To find: $P(X_6 = S_i|O;\theta)\ for\ i = 1,2$

This can be calculated using formula:

$$\gamma_t(i) = P(X_t = s_i|x_{1:T}) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i'}\ \alpha_t(i')\beta_t(i')}$$

Now,

$$\beta_6(i) = 1\ ; for\ all\ i$$

$$P(X_6 = s_1|O;\theta) = \frac{0.0002 * 1}{0.0002 * 1 +\ 0.000076 * 1}\ =\ 0.7246$$

$$P(X_6 = s_2|O;\theta) = \frac{0.000076 * 1}{0.0002 * 1 +\ 0.000076 * 1}\ =\ 0.2754$$

c) To find: $P(X_4 = S_i|O;\theta)\ for\ i = 1,2$

$$\beta_5(1) = 1\ *\ 0.8\ *\ 0.4\ +\ 1\ *\ 0.2\ *\ 0.2 =\ 0.36$$
$$\beta_5(2) = 1\ *\ 0.4\ *\ 0.4\ +\ 1\ *\ 0.6\ *\ 0.2 = 0.28$$

$$\beta_4(1) = 0.36\ *\ 0.8\ *\ 0.1\ +\ 0.28\ *\ 0.2\ *\ 0.3 = 0.0456$$
$$\beta_4(2) = 0.36\ \ *\ 0.4\ *\ 0.1\ +\ 0.28\ *\ 0.6\ *\ 0.3 = 0.0648$$

$$P(X_4 = s_1|O;\theta) = \frac{0.0043 * 0.0456}{0.0043 * 0.0456 +\ 0.0015 * 0.0648} = 0.6686$$

$$P(X_4 = s_2|O;\theta) = \frac{0.0015 * 0.0648}{0.0043 * 0.0456 +\ 0.0015 * 0.0648}\ =\ 0.3314$$

d) We want the most likely sequence of events given the current Output. We need to apply the viterbi algorithm for that.

$$\delta_t(j) = \max_{z_1,..,z_{t-1}} P\big(Z_1 = z_1,..,Z_t = s_j, x_{1:t}\big|\lambda\big)$$

The recusion eq is - $\delta_t(j) = \max_{\mathrm{i}} \delta_{t-1}(i)a_{ij}P(x_t|Z_t = s_j)$

The max probability is given by-

$$\arg\max_{\mathrm{j}} \delta_T(j)$$

Log delta values for S1-

t=1, -1.27

t=2, -2.41

t=2, -4.8

t=4, -6.1

t=5, -8.7

t=6, -9.8

For S2 –
t=1, -2.7
t=2, -4.5
t=2, -5.3
t=4, -7.34
t=5, -8.9
t=6, -11.0

Thus the most likely seq is 0
S1 -> S1 -> S1 -> S1 -> S1 -> S1

e) To find: $O_7 = \arg\max_{O} P(O|\boldsymbol{O};\theta)$

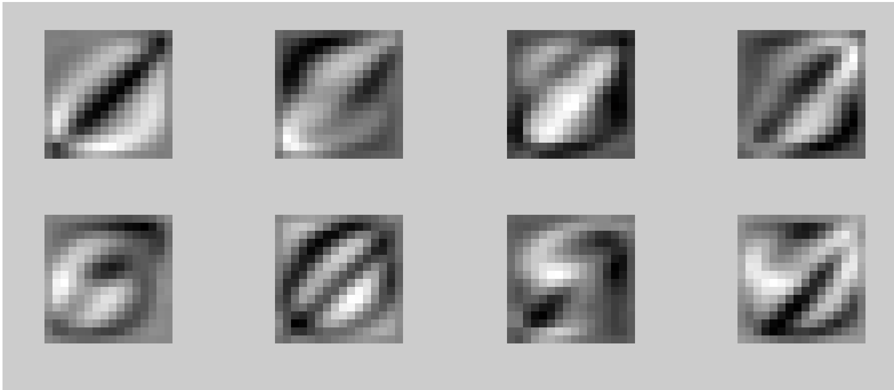
We need to predict the next sequence given the current sequence and parameters. We can also maximize on the joint likelihood.

$$P(O, \boldsymbol{O} = A|\theta) = \alpha_7(1) + \alpha_7(2) = 0.00008 + 0.00001 = 0.00009$$
$$P(O, \boldsymbol{O} = C|\theta) = \alpha_7(1) + \alpha_7(2) = 0.00002 + 0.00002 = 0.00004$$
$$P(O, \boldsymbol{O} = G|\theta) = \alpha_7(1) + \alpha_7(2) = 0.00008 + 0.00001 = 0.00009$$
$$P(O, \boldsymbol{O} = T|\theta) = \alpha_7(1) + \alpha_7(2) = 0.00002 + 0.00002 = 0.00004$$

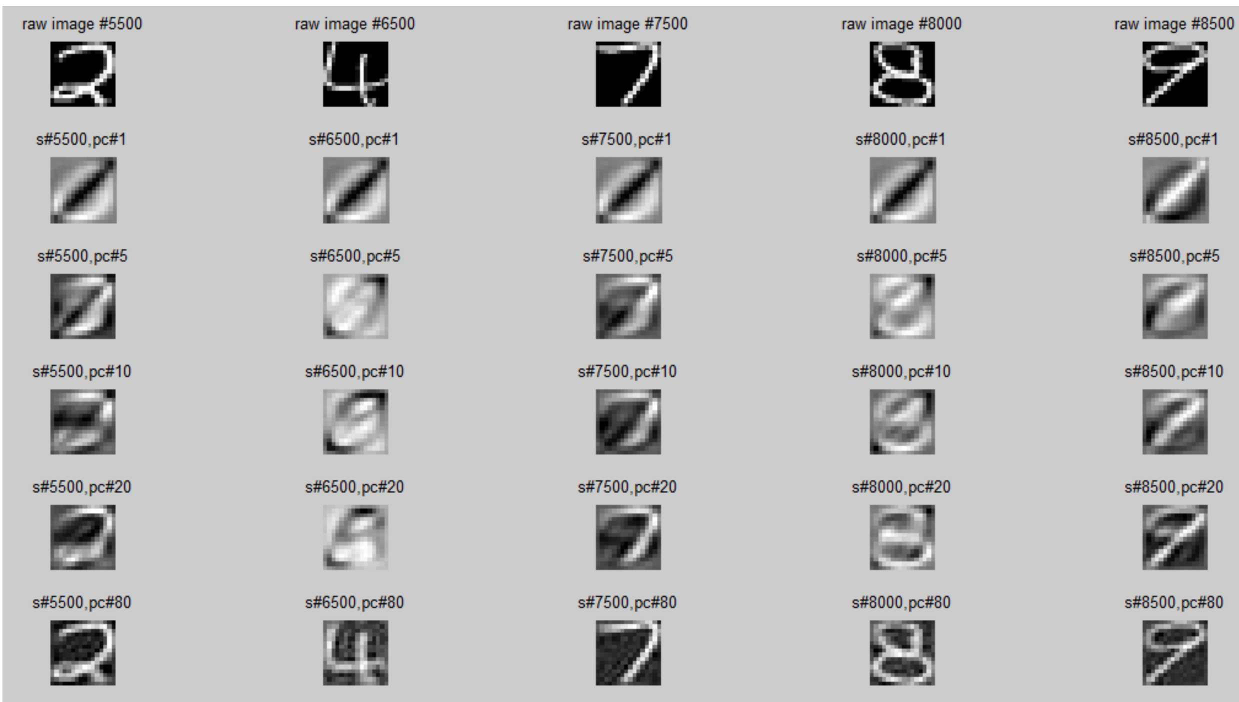
3. Programming

3.1 Principal Component Analysis.

a) Coded get sorted eigenvecs.m file.



b)



c)

d) As the K increases the accuracy increases as more eigen vectors are being used to represent the image and hence the classification is easier. The time also increases because the size of matrix being increases.

training accuracy for pc1 = 51.68
testing accuracy for pc1 = 19.20
time taken = 0.53 seconds

training accuracy for pc5 = 83.74
testing accuracy for pc5 = 50.30
time taken = 1.07 seconds
training accuracy for pc10 = 90.91
testing accuracy for pc10 = 72.23
time taken = 1.95 seconds
training accuracy for pc20 = 92.31
testing accuracy for pc20 = 74.60
time taken = 5.03 seconds
training accuracy for pc80 = 92.98
testing accuracy for pc80 = 76.70
time taken = 20.73 seconds

3.2 **Hidden Markov Model**:

3.2.1 Data Descriptions :Data was loaded via Matlab's import functionality.

3.2.2 Algorithm

3.2.3 Questions:

a) Length of shortest trace -109 , length of longest trace – 435, number of observations- 22

b) And c)

We need to find the log-likelyhood of training data given hidden state for all time series.
Observations- $X^1 .. X^D$
Underlying States- $Z = 1..M$
Standard HMM params- π, A, B

Now using the HMM assumptions (markov chain, independence), we can write the joint likelihood as:

$$P(X, Z|\theta) = \prod_{d=1}^{D} \left(\pi_{z_1^d} B_{z1^d}(x_1^d) \prod_{t=2}^{T} A_{z_{t-1}^d z_t^d} B_{z_t^d}(x_t^d) \right)$$

Taking log of this we'll get the log-likelyhood:

$$\log P(X, Z|\theta)$$

$$= \sum_{d=1}^{D} [\log \pi_{z_1^d} + \sum_{t=1}^{T} \log B_{z_t^d}(x_t^d) + \sum_{t=2}^{T} \log A_{z_{t-1}^d z_t^d}]$$

Now the EM steps are:

$$E\ step\colon\ Q(\theta, \theta^S) = \sum_{z \in Z} \log P(X, Z|\theta)\, P(z|X, \theta^s)$$

$$M\ step\colon\ \theta^{s+1} = \arg\max_{\theta} Q(\theta, \theta^S)$$

Putting the value of log likelihood calculated in b), in this equation.

$$Q(\theta, \theta^S) = \sum \sum_{d=1}^{D} \log \pi_{z_1^d} P(z|X, \theta^s)$$
$$+ \sum \sum_{d=1}^{D} \sum_{t=1}^{T} \log A_{z_{t-1}^d z_t^d} P(z|X, \theta^s)$$
$$+ \sum \sum_{d=1}^{D} \sum_{t=1}^{T} \log B_{z_t^d}(x_t^d)\ \ P(z|X, \theta^s)$$

Now we have some probability constraints:

$$\sum_{i=1}^{M} \pi_i = 1\ , \sum_{j=1}^{M} A_{ij} = 1, \sum_{j=1}^{N} B_i(j) = 1$$

We can thus optimize the Q function along with these constraints using the Lagrange multipliers.

$$L = Q(\theta, \theta^S) - \lambda_\pi \sum_{i=1}^{M} \pi_i - 1 - \sum_{i=1}^{M} \lambda_{A_i} (\sum_{j=1}^{M} A_{ij} - 1) -$$
$$\sum_{i=1}^{M} \lambda_{B_i} (\sum_{j=1}^{N} B_i(j) - 1)$$

Now deriving wrt. π_i:

$$\frac{\delta L}{\delta \pi_i} = \sum \frac{P\left(z_1^d = i, X|\theta^S\right)}{\pi_i} - \lambda_{pi} = 0$$

$$\frac{\delta L}{\delta \lambda_\pi} = -(\sum_{i=1}^{M} \pi_i - 1) = 0$$

On solving this further by marginalizing,

$$\pi_i = \frac{1}{D} \sum_{d=1}^{D} P(z_1(d) = 1|X^d, \theta^S)$$

Similarly for A_{ij}:

$$\frac{\delta L}{\delta A_{ij}} = \sum_{d=1}^{D}\sum_{t=2}^{T}\frac{P\left(z_{t-1}^{d} = i, z_{t}^{d} = j, X|\theta^{S}\right)}{A_{ij}} - \lambda_{A_i} = 0$$

$$\frac{\delta L}{\delta \lambda_{A_i}} = -\sum_{j=1}^{M} A_{ij} - 1 = 0$$

Solving this,

$$A_{ij} = \frac{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i, z_{t}^{d} = j|X^{d}, \theta^{S}\right)}{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i|X^{d}, \theta^{S}\right)}$$

Next, for $B_i(j)$:

$$\frac{\delta L}{\delta B_i(j)} = \sum_{d=1}^{D}\sum_{t=2}^{T}\frac{P\left(z_{t}^{d} = i, X|\theta^{S}\right)I(x_{t}^{d} = j)}{B_i(j)} - \lambda_{B_i} = 0$$

$$\frac{\delta L}{\delta \lambda_{B_i}} = -\sum_{j=1}^{N} B_i(j) - 1 = 0$$

$$B_i(j) = \frac{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i|X^{d}, \theta^{S}\right)I(x_{t}^{d} = j)}{\sum_{d=1}^{D}\sum_{t=1}^{T} P\left(z_{t}^{d} = i|X^{d}, \theta^{S}\right)}$$

So the updated summary steps are:

$$\pi_i^{s+1} = \frac{1}{D}\sum_{d=1}^{D} P(z_1^{d} = i|X^{d}, \theta^{S})$$

$$A_{ij}^{s+1} = \frac{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i, z_{t}^{d} = j|X^{d}, \theta^{S}\right)}{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i|X^{d}, \theta^{S}\right)}$$

$$B_i^{s+1}(j) = \frac{\sum_{d=1}^{D}\sum_{t=2}^{T} P\left(z_{t-1}^{d} = i|X^{d}, \theta^{S}\right)I(x_{t}^{d} = j)}{\sum_{d=1}^{D}\sum_{t=1}^{T} P\left(z_{t}^{d} = i|X^{d}, \theta^{S}\right)}$$