CSCI 567 Assignment 6
Fall 2016

Snehal Adsule
2080872073
adsule@usc.edu

November 22,2016

1 1 PCA

1.1 1.1 (a)

$$J = \frac{1}{N}\sum_{i=1}^{N}(x_i - p_{i1}e_i - p_{i2}e_2)^T(x_i - p_{i1}e_i - p_{i2}e_2)$$
$$= \frac{1}{N}\sum_{i=1}^{N}(x_i^2 - 2p_{i1}e_1x_i - 2p_{i2}e_2x_i - 2p_{i1}p_{i2}e_1^Te_2 + p_{i1}^2e_1^Te_1 + p_{i2}^2e_2^Te_2)$$

Given that $||e1||_2 = 1, ||e2||_2 = 1, and e_1^Te_2 = 0,$

$$J = \frac{1}{N}\sum_{i=1}^{N}(x_i^2 - 2p_{i1}e_1^Tx_i - 2p_{i2}e_2^Tx_i - 0 + p_{i1}^2.1 + p_{i2}^2.1)$$
$$\frac{\partial J}{\partial p_{i2}} = \frac{1}{N}\sum_{i=1}^{N}(0 - 0 - 2e_2^Tx_i + 0 + 2p_{i2})$$
$$= \frac{1}{N}\sum_{i=1}^{N}(-2e_2^Tx_i + 2p_{i2})$$

Setting to zero

$$\frac{1}{N}\sum_{i=1}^{N}(-2e_2^Tx_i + 2p_{i2}) = 0$$
$$\implies p_{i2} = e_2^Tx_i \ \forall \text{ i}$$

1.2 1.1 (b)

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

Setting derivative to zero

$$\begin{aligned}
&\frac{\partial \tilde{J}}{\partial e_2} = 0 \\
&\qquad \implies -2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 = 0 \quad \text{we multiply with } e_1^T \\
&\qquad \implies -2e_1^T S e_2 + 2\lambda_2 e_1^T e_2 + \lambda_{12} e_1^T e_1 = 0 \\
&\qquad \implies -2(Se_1)^T e_2 + 2\lambda_2 \times 0 + \lambda_{12} \times 1 = 0 \text{ since } S = S^T \\
&\qquad \text{As } (Se_1)^T e_2 = 0 \implies \lambda_{12} = 0 \\
&\qquad \implies Se_2 = \lambda_2 e_2 \text{ by substituting value of } \lambda_{12}
\end{aligned}$$

Since, $Se_2 = \lambda_2 e_2$, e_2 is the normalized eigenvector associated with the second largest eigenvector which minimizes $\tilde{J}$.

1.3 1.2 (a) Real Example

Used the following script to get the eigen values :

from numpy import linalg as LA
A=[[91.43 ,171.92, 297.99], [171.92, 373.92 ,545.21], [297.99,545.21,1297.26]]
w,v = LA.eig(A)
print w
for i in range(3):
print 'v',(i+1), '=', v[:,i]

The eigen values are [1626.52644399 , 7.09745924, 128.98609676]
and the three eigen vectors are :
v 1 = [0.21793758, 0.41449518, 0.88357057]
v 2 = [0.94428286, -0.31834854, -0.0835709]
v 3 = [-0.24664366, -0.85255378, 0.46078081]

1.4 1.2 (b)

We can observe that, λ_1 , λ_3 and λ_2 contribute to 92.27% , 7.3175% and 0.40267% of the variance of the bird data. The contribution of λ_2 is insignificant compared the other two, and can be neglected without much loss of the information after projection of points on the plane formed by the orthogonal vectors v_1 and v_3.

1.5 1.2 (c)

There are two main directions v_1 and $v_3$which captures the birds size with respect to length, wingspan, and weight. It is intutive that all the weight of the vectors are positive and therefore, shows direct relation between larger length, wingspan, and weight. Consider first principal component for v_1 and we see that feature weight has most impact on the bird's size as corresponsing weight 0.88 is larger than other two features.The $v3$ is the second largest component which is mainly dominated by the wingspan and the length but are in opposite direction with respect to corresponding weight value.

2 2.Hidden Markov Model

2.1 2 (a)

For the gene sequence $O = ACCGTA$, $P(O;\theta) = \sum_{j=1}^{2} \alpha_6(j)$
Base case:
$\alpha_1(j) = P(O_1|S_1 = j)P(S_1 = j)$ and
$\alpha_t(j) = P(O_t|S_t = j) \sum_{i=1}^{2} a_{ij}\alpha_{t-1}(j)$ otherwise $(\forall i > 1)$

$$
\begin{aligned}
&\alpha_1(1) = P(S_1 = 1)P(O_1|S_1 = 1) = \pi_1 \times P(A|1) = 0.6 \times 0.4 = 0.24 \\
&\alpha_1(2) = \pi_2 \times P(A|2) = 0.4 \times 0.2 = 0.08 \\
&\alpha_2(1) = P(O_2|S_2 = 1) \times \sum_i a_{i1}\alpha_1(j) \\
&= b_{1g} \times (a_{11}\alpha_1(1) + a_{21}\alpha_1(2)) = 0.04 \\
&\alpha_2(2) = b_{2c} \times (a_{12}\alpha_1(1) + a_{22}\alpha_1(2)) = 0.048 \\
&\alpha_3(1) = b_{1c} \times (a_{11}\alpha_2(1) + a_{21}\alpha_2(2)) = 0.00944 \\
&\alpha_3(2) = b_{2c} \times (a_{12}\alpha_2(1) + a_{22}\alpha_2(2)) = 0.01632 \\
&\alpha_4(1) = b_{1g} \times (a_{11}\alpha_3(1) + a_{21}\alpha_3(2)) = 0.0039408 \\
&\alpha_4(2) = b_{2g} \times (a_{12}\alpha_3(1) + a_{22}\alpha_3(2)) = 0.0012624 \\
&\alpha_5(1) = b_{1t} \times (a_{11}\alpha_4(1) + a_{21}\alpha_4(2)) = 0.000326352 \\
&\alpha_5(2) = b_{2t} \times (a_{12}\alpha_4(1) + a_{22}\alpha_4(2)) = 0.000581904 \\
&\alpha_6(1) = b_{1a} \times (a_{11}\alpha_5(1) + a_{21}\alpha_5(2)) = 0.000184483 \\
&\alpha_6(2) = b_{2a} \times (a_{12}\alpha_5(1) + a_{22}\alpha_5(2)) = 8.94096E-05
\end{aligned}
$$

$P(O;\theta) = \alpha_6(1) + \alpha_6(2) = 0.000273893$

2.2 2 (b)

$$
\begin{aligned}
\beta_{t-1}(i) = \sum_{j=1}^{2} \beta_t a_{ij} P(O_t|X_t = S_j) \\
\beta_6(1) &= 1 \\
\beta_6(2) &= 1 \\
\beta_5(1) &= \beta_6(1)a_{11}b_{1a} + \beta_6(2)a_{12}b_2 = 0.28 \\
\beta_5(2) &= \beta_6(1)a_{21}b_{1a} + \beta_6(2)a_{22}b_{2a} = 0.34 \\
\beta_4(1) &= \beta_5(1)a_{11}b_{1t} + \beta_5(2)a_{12}b_2 = 0.064 \\
\beta_4(2) &= \beta_5(1)a_{21}b_{1t} + \beta_5(2)a_{22}b_{2a} = 0.049
\end{aligned}
$$

$$
\begin{aligned}
P(X_6 = S_1|O,\theta) &= \frac{\alpha_6(S_1)\beta_6(S_1)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} \\
&= \frac{0.000184483 \times 1}{0.000184483 \times 1 + 8.94096E - 05 \times 1} \\
&= 0.673559875 \\
P(X_6 = S_2|O,\theta) &= \frac{\alpha_6(S_2)\beta_6(S_2)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} \\
&= \frac{8.94096E - 05 \times 1}{0.000184483 \times 1 + 8.94096E - 05 \times 1} \\
&= 0.326440125
\end{aligned}
$$

2.3 2 (c)

Similarly,

$$
\begin{aligned}
P(X_4 = S_1|O,\theta) &= \frac{\alpha_4(S_1)\beta_4(S_1)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} \\
&= \frac{0.0039408 \times 0.064}{0.0039408 \times 0.064 + 0.0012624 \times 0.049} \\
&= 0.705017437 \\
P(X_4 = S_2|O,\theta) &= \frac{\alpha_4(S_2)\beta_4(S_2)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} \\
&= \frac{0.0012624 \times 0.049}{0.0039408 \times 0.064 + 0.0012624 \times 0.049} \\
&= 0.294982563
\end{aligned}
$$

2.4 2 (d)

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

$$
\begin{aligned}
\delta_1(1) =& \pi_1 b_{1a} = 0.24 \\
\delta_1(2) =& \pi_2 b_{2a} = 0.08 \\
\delta_2(1) =& b_{1c} \times max(\delta_1(1)a_{11}, \delta_1(2)a_{21}) = 0.0336 \\
\delta_2(2) =& b_{2c} \times max(\delta_1(1)a_{12}, \delta_1(2)a_{22}) = 0.0288 \\
\delta_3(1) =& b_{1c} \times max(\delta_2(1)a_{11} + \delta_2(2)a_{21}) = 0.004704 \\
\delta_3(2) =& b_{1c} \times max(\delta_2(1)a_{11} + \delta_2(2)a_{21}) = 0.006912 \\
\delta_4(1) =& b_{1g} \times max(\delta_3(1)a_{11} + \delta_3(2)a_{21}) = 0.00098784 \\
\delta_4(2) =& b_{1g} \times max(\delta_3(1)a_{11} + \delta_3(2)a_{21}) = 0.00041472 \\
\delta_5(1) =& b_{1t} \times max(\delta_4(1)a_{11} + \delta_4(2)a_{21}) = 0.0000691 \\
\delta_5(2) =& b_{1t} \times max(\delta_4(1)a_{11} + \delta_4(2)a_{21}) = 0.0000889 \\
\delta_6(1) =& b_{1a} \times max(\delta_5(1)a_{11} + \delta_5(2)a_{21}) = 0.0000194 \\
\delta_6(2) =& b_{1a} \times max(\delta_5(1)a_{11} + \delta_5(2)a_{21}) = 0.0000107
\end{aligned}
$$

Most likely path = arg $max_j \delta_T(j)$ =s1,s1,s2,s1,s2,s1

2.5 2 (e)

Lets assume the $O_7 = x$,where $x \in (A, C, T, G)$

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

$$
\begin{aligned}
P(O_7 = A|\theta) =& 0.4 \times (0.673559875 \times 0.7 + 0.326440125 \times 0.4) \\
&+ 0.2 \times (0.673559875 \times 0.4 + 0.326440125 \times 0.6) = 0.33388479 \\
P(O_7 = T|\theta) =& 0.2 \times (0.673559875 \times 0.7 + 0.326440125 \times 0.4) \\
&+ 0.4 \times (0.673559875 \times 0.4 + 0.326440125 \times 0.6) = 0.306528803 \\
P(O_7 = C|\theta) =& 0.3 \times (0.673559875 \times 0.7 + 0.326440125 \times 0.4) \\
&+ 0.1 \times (0.673559875 \times 0.4 + 0.326440125 \times 0.6) = 0.227149191 \\
P(O_7 = G|\theta) =& 0.1 \times (0.673559875 \times 0.7 + 0.326440125 \times 0.4) \\
&+ 0.3 \times (0.673559875 \times 0.4 + 0.326440125 \times 0.6) = 0.199793204
\end{aligned}
$$

We observe that the observation A is more probable than the others, $P(O_7 = A|O_{1:6})$.