Name : Achyut Kulkarni

1 Principal Component Analysis

Given $e_1{}^T e_2 = 0$ *and* $\|e_1\|_2 = 1$ *and* $\|e_2\|_2 = 1$ $e_1 e_2{}^T = 0$

a) Given the cost function J

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

$$\frac{\delta J}{\delta p_{i2}} = \frac{1}{N}\sum_{i=1}^{N}(-e_2{}^T)(x_i - p_{i1}e_1 - p_{i2}e_2) + (x_i - p_{i1}e_1 - p_{i2}e_2)^T(-e_2) = 0$$

$$=> \sum_{i=1}^{N} - e_2{}^T x_i + p_{i1}e_1 e_2{}^T + p_{i2}e_2 e_2{}^T - e_2(x_i)^T + (p_{i1}e_1)^T e_2 + (p_{i2}e_2)^T e_2 = 0$$

$$=> \sum_{i=1}^{N} - e_2{}^T x_i + p_{i2} + p_{i2} - e_2{}^T x_i = 0$$

Since e and x are both vectors

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

$$=> \sum_{i=1}^{N} - 2e_2{}^T x_i + 2p_{i2} = 0$$

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

b) Given the cost function

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

To find the e_2 value that minimises the cost function, we find $\frac{\delta \widehat{J}}{\delta e_2}$ and equate to 0

$$\frac{\delta \widehat{J}}{\delta e_2} = -(S + S^T)e_2 + 2\lambda_2 e_2 + \lambda_{12}e_1 = 0$$

Because it is given that $\frac{\delta y^T A y}{\delta y} = (A + A^T)y$

S is symmetric

Hence,

$$\frac{\delta \widehat{J}}{\delta e_2} = - 2Se_2 + 2\lambda_2 e_2 + \lambda_{12}e_1$$

Pre-multiplying with $e_1{}^T$

$$\frac{\delta \widehat{J}}{\delta e_2} = -2e_1{}^T S e_2 + 2e_1{}^T \lambda_2 e_2 + e_1{}^T \lambda_{12} e_1$$

$$=> -2(S^T e_1)^T e_2 + 2\lambda_2 e_1{}^T e_2 + \lambda_{12} e_1{}^T e_1$$

Since S is symmetric and λ_1 is Eigen vector

$$Se_1 = \lambda e_1$$

$$=> -2\lambda_1 e_1{}^T e_2 + 2\lambda_2 e_1{}^T e_2 + \lambda_{12} e_1{}^T e_1 = 0$$

$=> \lambda_{12} = 0$

Substituting in $\frac{\delta \hat{J}}{\delta e_2}$

$$-2Se_2 + 2\lambda_2 e_2 + 0 * e_1 = 0$$

$$Se_2 = \lambda_2 e_2$$

Hence, the value of e_2 that minimizes cost function is given by the eigenvector associated with the second largest eigenvalue of S

1.2

a)

S = [91.43 171.92 297.99]
[373.92 545.21]
[1297.26]

The eigen value can be calculated by $det[s - \lambda I] = 0$

$$det \begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$=> det \begin{bmatrix} 91.43 - \lambda & 171.92 & 297.99 \\ 171.92 & 373.92 - \lambda & 545.21 \\ 297.99 & 545.21 & 1297.26 - \lambda \end{bmatrix}$$

=>(91.43-λ){(373.93-λ)(1297.26-λ)-(545.21)(545.21)}-171.92{(171.92)(1297.26-λ)-(297.99)(545.21)} + 297.99{(171.92)(545.21)-(297.99)(373.93-λ)}

=> Solving the cubic equations we get

$\lambda_1 = 1626.52 \quad \lambda_2 = 128.99 \quad \lambda_3 = 7.097$

For each eigen values, the corresponding Eigen vectors are:

$(S - \lambda I)V = 0$

Where V is eigen vector and can be written as

$[x_1]$
$[x_2]$
$[x_3]$

For λ_1 :

$$\begin{bmatrix} -1535.096 & 171.92 & 297.99 \\ 171.92 & 1252.606 & 545.21 \\ 297.9 & 545.21 & 329.266 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

x_1 =0.22 x_2 = 0.41 x_3 = 0.88

For λ_2 :

[-37.556 171.92 297.99] [x_1]

[171.92x 244.934 545.21] [x_2] = 0

[297.99 545.211168.274] [x_3]

x_1 =0.944 x_2 =-0.318 x_3 =-0.084

For λ_3 :

[84.333 171.92 297.99]

[171.92 366.823 545.21] = 0

[297.99 545.21 1290.1634]

x_1 =-0.247 x_2 =-0.853 x_3 =0.461

Eigen vectors are:

V_1 = 0.218 V_2 =-0.247 V_3 = 0.944

0.414 0.853 -0.318

0.884 -0.461 -0.084

b)

The contribution can be calculated as

$$\frac{\lambda}{sum(\lambda)} *100$$

$$c_1 = \frac{1626.52}{1626.52+128.99+7.097} = 92.3$$

$$c_2 = \frac{128.99}{1626.52+128.99+7.097} = 7.3$$

$$c_3 = \frac{7.097}{1626.52+128.99+7.097} = 0.4$$

Because Most of the information is contained in first two orthonormal directions. Hence the third (weight) can be dropped.

c)As done above, the information contained in each orthonormal directions can quantified. The length contains the most the information with 92.3%

2)

Observed sequence O = ACCGTA

To find the probability , we need to calculate the forward probability by formula

$$P(O_{1..T},\ \theta)\ = \sum_j \alpha_t(j)$$

Where $\alpha_t(j)\ =\ P(X_t =\ S_j;\ O_{1..t})$

Base condition can be given by:

$a_1(j)\ =\ P(O_1|X_1 = S_j)P(X_1 =\ S_j)\ =\ P(O_1|X_1 = S_j)\pi_j$,where π_j being the prior probability of the state S_j

$$\alpha_1(1) = P(O_1|X_1 = S_1)\pi_1 = 0.24$$
$$\alpha_1(2) = P(O_1|X_1 = S_2)\pi_2 = 0.08$$
$$\alpha_2(1)\ =\ b_{1C}(\alpha_1(1)a_{11}\ +\ \alpha_1(2)a_{21}) =\ 0.04$$
$$\alpha_2(2)\ =\ b_{2C}(\alpha_1(1)a_{12}\ +\ \alpha_1(2)a_{22}) = 0.048$$
$$\alpha_3(1)\ =\ b_{1C}(\alpha_2(1)a_{11} +\ \alpha_2(2)a_{21}) = 9.44e^{-3}$$
$$\alpha_3(2) =\ b_{1C}(\alpha_2(1)a_{12} +\ \alpha_2(2)a_{22}) = 0.01632$$
$$\alpha_4(1)\ =\ b_{1G}(\alpha_3(1)a_{11}\ + \alpha_3(2)a_{21}) = 3.940e^{-3}$$
$$\alpha_4(2)\ =\ b_{2G}(\alpha_3(1)a_{12}\ +\ \alpha_3(2)a_{22}) = 1.262e^{-3}$$
$$\alpha_5(1)\ =\ b_{1T}(\alpha_4(1)a_{11} + \alpha_4(2)a_{21}) = 3.263e^{-4}$$
$$\alpha_5(2)\ =\ b_{2T}(\alpha_4(1)a_{12} + \alpha_4(2)a_{22}) = 5.819e^{-4}$$
$$\alpha_6(1)\ =\ b_{1A}(\alpha_5(1)a_{11} + \alpha_5(2)a_{21}) = 1.844e^{-4}$$
$$\alpha_6(2)\ =\ b_{2A}(\alpha_5(1)a_{12} + \alpha_5(2)a_{22}) = 8.94e^{-5}$$

Probability of observed sequence $\alpha_6(1) + \alpha_6(2)\ =\ 1.844e^{-4} + 8.94e^{-5}\ =\ 2.738e^{-4}$

b) Base condition $\beta_6(1)\ = 1\ and\ \beta_6(2)\ =\ 1$

$$\beta_5(1)\ = \beta_6(1)a_{11}b_{1A} + \beta_6(2)a_{12}b_{2A} = 0.34$$
$$\beta_5(2)\ = \beta_6(1)a_{21}b_{1A} + \beta_6(2)a_{22}b_{2A} = 0.28$$
$$\beta_4(1)\ = \beta_5(1)a_{11}b_{1T} + \beta_5(2)a_{12}b_{2T} = 0.049$$
$$\beta_4(2)\ = \beta_5(1)a_{21}b_{1T} + \beta_5(2)a_{22}b_{2T} = 0.064$$

$$P(X_6 = S_1|O;\theta\,)\ =\ \frac{\alpha_6(1)\beta_t(1)}{\alpha_6(1)\beta_t(1)+\alpha_6(2)\beta_t(2)} = 0.67355$$

$$\sum_i P(X_j = i)\ =\ 1$$

$$P(X_6 = S_2|O;\theta\,) = 1 - P(X_6 = S_1|O;\theta\,)\ = 0.32645$$

c)

$$P(X_4 = S_1|O;\theta) = \frac{\alpha_4(1)\beta_4(1)}{\alpha_4(1)\beta_4(1)+\alpha_4(2)\beta_4(2)} = 0.7050$$
$$P(X_4 = S_2|O;\theta) = 1 - P(X_4 = S_1|O;\theta) = 0.2950$$

d)Viterbi algorithm gives the state chosen by

$$\delta_t(j) = max_i\,(\delta_{t-1}(i)\, a_{ij}\, P(O_t|X_t = S_j))$$

$\delta_1(1) = P(O_1|X_1 = S_1)\pi_1 = 0.24$

$\delta_1(2) = P(O_1|X_1 = S_2)\pi_2 = 0.08$

$\delta_2(1) = max\{b_{1C}\,\delta_1(1)a_{11}\,, b_{1C}\delta_1(2)a_{21})\} = 0.0336$

$\delta_2(1) = max\{\,b_{2C}\,\delta_1(1)a_{12}\,, b_{2C}\delta_1(2)a_{22})\} = 0.0288$

$\delta_3(1) = max\{b_{1C}\,\delta_2(1)a_{11}\,, b_{1C}\delta_2(2)a_{21})\} = 0.004704$

$\delta_3(2) = max\{\,b_{2C}\,\delta_2(1)a_{12}\,, b_{2C}\delta_2(2)a_{22})\} = 0.006912$

$\delta_4(1) = max\{b_{1G}\,\delta_3(1)a_{11}\,, b_{1G}\delta_3(2)a_{21})\} = 0.0009878$

$\delta_4(2) = max\{\,b_{2G}\,\delta_3(1)a_{12}\,, b_{2G}\delta_3(2)a_{22})\} = 0.0004147$

$\delta_5(1) = max\{b_{1T}\,\delta_4(1)a_{11}\,, b_{1T}\delta_4(2)a_{21})\} = 0.00006914$

$\delta_5(2) = max\{\,b_{2T}\,\delta_4(1)a_{12}\,, b_{2T}\delta_4(2)a_{22})\} = 0.00008890$

$\delta_6(1) = max\{b_{1A}\,\delta_5(1)a_{11}\,, b_{1A}\delta_5(2)a_{21})\} = 0.0000193$

$\delta_6(2) = max\{\,b_{2A}\,\delta_5(1)a_{12}\,, b_{2A}\delta_5(2)a_{22})\} = 0.00001066$

Most likely path $S1 \text{ –> } S1 \text{ –> } S1 \text{ –> } S1 \text{ –> } S1 \text{ –> } S1$

e)

Compute $P(O_7|O;\Theta)$. *Then which observation is most likely after* $o_{1:6}$?

$(O_7 = argmax_O P(O|O;\Theta))$

To find max ($P(O|O = A;\Theta)$, $P(O|O = C;\Theta), P(O|O = G;\Theta), P(O|O = T;\Theta)$)

Let $m_1 = P(X_6 = S_1|O;\Theta) \times a_{11} + P(X_6 = S_2|O;\Theta) \times a_{21}$

And $m_2 = P(X_6 = S_1|O;\Theta) \times a_{12} + P(X_6 = S_2|O;\Theta) \times a_{22}$

Therefore,

$m_1 = (0.673559 \times 0.7) + (0.32644012 \times 0.4) = 0.60206$

$m_2 = (0.673559 \times 0.3) + (0.32644012 \times 0.6) = 0.39793$

$$P(O|O = A;\Theta) = m_1 \times b_{1A} + m_2 \times b_{2A}$$
$$= 0.60206 \times 0.4 + 0.39793 \times 0.2$$

$$= 0.32041$$

$$P(O|O = C;\Theta) = m_1 \times b_{1C} + m_2 \times b_{2C}$$
$$= 0.60206 \times 0.2 + 0.39793 \times 0.4$$
$$= 0.279584$$

$$P(O|O = G;\Theta) = m_1 \times b_{1G} + m_2 \times b_{2G}$$
$$= 0.60206 \times 0.3 + 0.39793 \times 0.1$$
$$= 0.220411$$

$$P(O|O = T;\Theta) = m_1 \times b_{1T} + m_2 \times b_{2T}$$
$$= 0.60206 \times 0.1 + 0.39793 \times 0.3$$
$$= 0.179585$$

Most likely observation at state 7 is A .