1. Principal Component Analysis

1.1 Derivation of Second Principal Component

(a) Given the cost function,

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

where $e_1\ and\ e_2$ are orthonormal vector basis for dimensionality reduction, i.e. $||e_1|| = 1, ||e_2|| = 1$ and $e_1^T e_2 = 0$ and some coefficients $p_{i1}\ and\ p_{i2}$

Let's define $m_i = x_i - p_{i1}e_1$.

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

$$= \frac{1}{N}\sum_{i=1}^{N}(m_i^T m_i - p_{i2}m_i^T e_2 - p_{i2}e_2^T k_i + p_{i2}^2 e_2^T e_2)$$

$$\frac{\partial J}{\partial p_{i2}} = \frac{1}{N}\sum_{i=1}^{N}(0 - m_i^T e_2 - e_2^T k_i + 2p_{i2}e_2^T e_2)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(-2e_2^T m_i + 2p_{i2})$$

Putting,

$$\frac{\partial J}{\partial p_{i2}} = 0$$

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}(-2e_2^T m_i + 2p_{i2}) = 0$$

$$\Rightarrow p_{i2} = e_2^T m_i \forall i$$

$$\Rightarrow p_{i2} = e_2^T(x_i - p_{i1}e_1)$$

Since, $e_2^T e_1 = 0$, we get,

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

(b) Given the cost function,

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

λ_2 is the Lagrange Multiplier for equality constraint $e_2^T e_1 = 1$ and λ_{12} is the Lagrange Multiplier for equality constraint $e_2^T e_1 = 0$.

Using

$$\frac{\partial y^T A y}{\partial y} = (A + A^T)y$$

and substituting the Lagrange's multipliers for $e_2^T e_2$ and $e_2^T e_1$ we get,

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

Since $S = S^T$,

$$= -2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1$$

For minimizing w.r.t. e_2

$$\frac{\partial \tilde{J}}{\partial e_2} = 0$$

$$\Rightarrow -2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 = 0$$

Multiplying the equation by e_1^T

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

Since $S = S^T$,

$$\Rightarrow -2(Se_1)^T e_2 + 2\lambda_2 \, X \, 0 + \lambda_{12} \, X \, 1 = 0$$

Since $(Se_1)^T e_2 = 0$

$$\lambda_{12} = 0$$

$$\Rightarrow Se_2 = \lambda_2 e_2$$

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

$$\boxed{Se_2 = \lambda_2 e_2}$$

1.2 A Real Example

Given,

$$S = \begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix}$$

Dividing all the elements by n-1 i.e. 99, we get,

$$S = \frac{1}{99}\begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix}$$

(a) Finding eigenvalues,

$$|S - \lambda I| = 0$$

$$S - I\lambda = \frac{1}{99}\begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix} - \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{99}\begin{bmatrix} 91.43 - \lambda_1 & 171.92 & 297.99 \\ 171.92 & 373.92 - \lambda_2 & 545.21 \\ 297.99 & 545.21 & 1297.26 - \lambda_3 \end{bmatrix}$$

Therefore,

$$\Rightarrow \frac{1}{99}\begin{vmatrix} 91.43 - \lambda_1 & 171.92 & 297.99 \\ 171.92 & 373.92 - \lambda_2 & 545.21 \\ 297.99 & 545.21 & 1297.26 - \lambda_3 \end{vmatrix} = 0$$

$$\frac{1}{99}\Big[(91.43 - \lambda_1)\big((373.92 - \lambda_2)(1297.26 - \lambda_3) - 545.21^2\big) + 171.92\,\Big(\big(171.92 * (1297.26 - \lambda_3)\big) - \big((373.92 - \lambda_2) * 297.99\big)\Big) + 297.99((171.92 * 545.21) - \big((373.92 - \lambda_2) * 297.99)\big)\Big] = 0$$

Solving the above equation, we get,

$$\lambda_1 = 1626.5$$

$$\lambda_2 = 129$$

$$\lambda_3 = 7.1$$

Using the eigenvalues, we calculate eigenvectors using,

$$SX = \lambda X$$

$$(S - \lambda)X = 0$$

We get three equations for the three values of λ, can be written as,

$$(S - 1626.5 \ I)X_1 = 0$$

$$(S - 129 \ I)X_2 = 0$$

$$(S - 7.1 \ I)X_3 = 0$$

Solving the above equations, we get,

$$\mathrm{X}_1 = \begin{bmatrix} 0.22 \\ 0.41 \\ 0.88 \end{bmatrix}$$

$$\mathrm{X}_2 = \begin{bmatrix} 0.25 \\ 0.85 \\ -0.46 \end{bmatrix}$$

$$\mathrm{X}_3 = \begin{bmatrix} 0.94 \\ -0.32 \\ -0.08 \end{bmatrix}$$

(b)

We see that the value of λ_1 is far greater than that of λ_2 and λ_3. Here,

The principal component from X_1 constitutes for below % variation in data,

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{1626.5}{1626.5 + 129 + 7.1} \approx 92.28\ \%$$

The principal component from X_2 constitutes for below % variation in data,

$$\frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{129}{1626.5 + 129 + 7.1} \approx 7.32\%$$

The principal component from X_3 constitutes for below % variation in data,

$$\frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{7.1}{1626.5 + 129 + 7.1} \approx 0.40\%$$

From this we can see that λ_3 accounts to negligible variance of the data, hence we can omit the eigenvector X_3.

(c)

Since we've decided that the vector X_3, we'll interpret X_1 and X_2.

The values in X_1 (first principal component) all have a positive value, which indicate that the birds with larger size tend to have larger length, wingspan and weight. In this vector the third value corresponding to the weight is the highest value, it's change is influences the bird size the most followed by the change in the second value corresponding to the wingspan and finally the length which affects the size by the least factor.

The values in X_2 (second principal component) have the highest values in the second and third entries when we consider the absolute values and they affect the size of the birds of the most. Here, we see that the values of second and third entries are of opposite sign indicating that the birds with larger wingspans have smaller weights and vice-versa. The length of the bird here has the least effect on the size of the bird.

2. Hidden Markov Model

(a)

$$a = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Given, O = ACCGTA,

$$P(O;\Theta) = \sum_{j=1}^{2} \alpha_6(j)$$

where $\alpha_t(j) = P(O_t|S_t = j)\sum_{i=1}^{2} a_{ij}\alpha_{t-1}(j)$

$\alpha_1(j) = P(O_1|S_1 = j)P(S_1 = j)$

Also, $for\ i\ >\ 1, \alpha_t(j) = P(O_t|S_t = j)\ X\ \sum_i a_{ij}\alpha_{t-1}(j)$

Therefore,

$b_{1A} = 0.4,\ b_{1C} = 0.2, b_{1G} = 0.3, b_{1T} = 0.1, b_{2A} = 0.2, b_{2C} = 0.4, b_{2G} = 0.1, b_{2T} = 0.3$

Substituting these values in the below equations, we get,

$$\alpha_1(1) = P(O_1|S_1 = 1)\, P(S_1 = 1) = b_{1A} \times \pi_1 = 0.4 \times 0.6 = 0.24$$

and $\alpha_1(2) = b_{2A} \times \pi_2 = 0.2 \times 0.4 = 0.08$

$$\alpha_2(1) = P(O_2|S_2 = 1) \times \sum_i a_{i1}\alpha_1(j) = b_{1C} \times \left(a_{11}\alpha_1(1) + a_{21}\alpha_1(2)\right) = 0.04$$

$$\alpha_2(2) = b_{2C} \times \left(a_{12}\alpha_1(1) + a_{22}\alpha_1(2)\right) = 0.048$$

$$\alpha_3(1) = b_{1C} \times \left(a_{11}\alpha_2(1) + a_{21}\alpha_2(2)\right) = 9.44e^{-3}$$

$$\alpha_3(2) = b_{2c} \times \left(a_{12}\alpha_2(1) + a_{22}\alpha_2(2)\right) = 1.632e^{-2}$$

$$\alpha_4(1) = b_{1g} \times \left(a_{11}\alpha_3(1) + a_{21}\alpha_3(2)\right) = 3.94e^{-3}$$

$$\alpha_4(2) = b_{2g} \times \left(a_{12}\alpha_3(1) + a_{22}\alpha_3(2)\right) = 0.001262$$

$$\alpha_5(1) = b_{1t} \times \left(a_{11}\alpha_4(1) + a_{21}\alpha_4(2)\right) = 0.00032$$

$$\alpha_5(2) = b_{2t} \times \left(a_{12}\alpha_4(1) + a_{22}\alpha_2(2)\right) = 5.81e^{-4}$$

$$\alpha_6(1) = b_{1a} \times \left(a_{11}\alpha_5(1) + a_{21}5(2)\right) = 1.84e^{-4}$$

$$\alpha_6(2) = b_{2a} \times \left(a_{12}\alpha_5(1) + a_{22}\alpha_5(2)\right) = 8.94e^{-5}$$

Therefore, the total probability of the sequence,

$P(O;\Theta\,) = \alpha_6(1) + \alpha_6(2) = 1.844e^{-4} + 8.94e^{-5} = 2.738\ e^{-4}$

(b)

Let the hidden states $S_1 and\ S_2$ be represented as 1 and 2 respectively.

$$\beta_{t-1}(i) = \sum_{j=1}^{2} \beta_t a_{ij} P(O_t|X_t = S_j)$$

Base conditions, $\beta_6(1) = 1\ and\ \beta_6(2) = 1$

$$\beta_5(1)\ = \beta_6(1)b_{1A}a_{11} + \beta_6(2)b_{2A}a_{12} = 0.34$$

$$\beta_4(1)\ = \beta_5(1)b_{1T}a_{11} + \beta_5(2)b_{2T}a_{12} = 4.9e^{-2}$$

$$\beta_5(2)\ = \beta_6(1)b_{1A}a_{21} + \beta_6(2)b_{2A}a_{22} = 0.28$$

$$\beta_4(2)\ = \beta_5(1)b_{1T}a_{21} + \beta_5(2)b_{2T}a_{22} = 6.4e^{-2}$$

$$P(X_6 = S_1|O;\Theta)\ = \frac{\alpha_6(1)\beta_6(1)}{\alpha_6(1)\beta_6(1) + \alpha_6(2)\beta_6(2)} = 0.67355$$

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

(c)

$$P(X_4 = S_1|O;\Theta) = \frac{\alpha_4(1)\beta_4(1)}{\alpha_4(1)\beta_4(1) + \alpha_4(2)\beta_4(2)} = \frac{0.00019306}{0.00027382} = 0.7050$$

$$P(X_4 = S_2|O;\Theta)\ = 1 - P(X_4 = S_1|O;\Theta)\ = 0.2950$$

(d)

From Viterbi algorithm we get the state chosen by

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

Now we have, $\delta_1(1) = P(O_1|X_1 = S_1)\pi_1 = 0.24$

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

Similarly, computing it for the other states, we get,

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

$\delta_3(1) = \max[\delta_2(1)b_{1C}a_{11}, \delta_2(2)b_{1C}a_{21})] = 4.704e^{-3}$

$\delta_4(1) = \max[\delta_3(1)b_{1G}a_{11}, \delta_3(2)b_{1G}a_{21})] = 9.878\ e^{-4}$

$\delta_5(1) = max[\delta_4(1)b_{1T}a_{11}, \delta_4(2)b_{1T}a_{21})] = 6.914e^{-5}$

$\delta_6(1) = max[\delta_5(1)b_{1A}a_{11}, \delta_5(2)b_{1A}a_{21})] = 1.93e^{-5}$

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

$\delta_3(2) = \max[\ \delta_2(1)b_{2C}a_{12}\ , \delta_2(2)b_{2C}a_{22})] = 6.912e^{-3}$

$\delta_4(2) = max[\ \delta_3(1)b_{2G}a_{12}\ , \delta_3(2)b_{2G}a_{22})] = 4.147e^{-4}$

$\delta_5(2) = max[\delta_4(1)\ b_{2T}a_{12}\ , \delta_4(2)\ b_{2T}a_{22})] = 8.890e^{-5}$

$\delta_6(2) = max[\ \delta_5(1)b_{2A}a_{12}\ , \delta_5(2)b_{2A}a_{22})] = 1.066e^{-5}$

(e)

From the previously calculated values,

$$P(X_6 = S_1|O;\Theta) = \mathbf{0.67355}$$

$$P(X_6 = S_2|O;\Theta) = \mathbf{0.32645}$$

For O_7=A:

$$P(A|O;\Theta) = P(X_7 = S_1|\ O;\Theta) + P(X_7 = S_2|\ O;\Theta)$$

$$= \{P(X_6 = S_1|O;\Theta)b_{1A}a_{11} + P(X_6 = S_2|O;\Theta)b_{1A}a_{21}\} + \{P(X_6 = S_1|O;\Theta)b_{2A}a_{12} + P(X_6 = S_2|O;\Theta)b_{2A}a_{22}\} = \mathbf{0.3204}$$

For O_7=C:

$$P(C|O;\Theta) = P(X_7 = S_1|\ O;\Theta) + P(X_7 = S_2|\ O;\Theta)$$

$$= \{P(X_6 = S_1|O;\Theta)b_{1C}a_{11} + P(X_6 = S_2|O;\Theta)b_{1C}a_{21}\} + \{P(X_6 = S_1|O;\Theta)b_{2C}a_{12} + P(X_6 = S_2|O;\Theta)b_{2C}a_{22}\} = \mathbf{0.2795}$$

For O_7=G:

$$P(G|O;\Theta) = P(X_7 = S_1|\ O;\Theta) + P(X_7 = S_2|\ O;\Theta)$$

$$= \{P(X_6 = S_1|O;\Theta)b_{1G}a_{11} + P(X_6 = S_2|O;\Theta)b_{1G}a_{21}\} + \{P(X_6 = S_1|O;\Theta)b_{2G}a_{12} + P(X_6 = S_2|O;\Theta)b_{2G}a_{22}\} = \mathbf{0.2204}$$

For O_7=T:

$$P(T|O;\Theta) = P(X_7 = S_1|\ O;\Theta)\ +\ P(X_7 = S_2|\ O;\Theta)$$

$$= \{P(X_6 = S_1|O;\Theta)b_{1T}a_{11}\ + P(X_6 = S_2|O;\Theta)b_{1T}a_{21}\} \\ +\ \{P(X_6 = S_1|O;\Theta)b_{2T}a_{12}\ +\ P(X_6 = S_2|O;\Theta)b_{2T}a_{22}\} = \mathbf{0.1795}$$

Most likely observation predicted as 7th observed sequence (after $O_{1:6}$) is A.

COLLABORATION

Brain stormed and Collaborated with Adarsha Desai and Ravishankar Sivaraman for this assignment.