1. Principal Component Analysis

1.1 Derivation of Second Principal Component

(a).

We are given,

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

Let's perform $\frac{\partial J}{\partial p_{i2}} = 0$

$$\frac{\partial}{\partial p_{i2}} \left(\frac{1}{N} \sum_{i=1}^{N} (x_i - p_{i1} e_i - p_{i2} e_2)^T (x_i - p_{i1} e_i - p_{i2} e_2) \right) = 0$$

To simplify the computations, lets substitute $\alpha_i = x_i - p_{i1} e_1$

$$\frac{\partial}{\partial p_{i2}} \left(\frac{1}{N} \sum_{i=1}^{N} (\alpha_i - p_{i2} e_2)^T (\alpha_i - p_{i2} e_2) \right) = 0$$

$$\frac{\partial}{\partial p_{i2}} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\alpha_i^T \alpha_i - 2\, p_{i2} e_2^T \alpha_i + p_{i2}^2 e_2^T e_2 \right) \right) = 0$$

$$\frac{\partial}{\partial p_{i2}} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\alpha_i^T \alpha_i - 2\, p_{i2} e_2^T \alpha_i + p_{i2}^2 e_2^T e_2 \right) \right) = 0$$

$$\left(\frac{1}{N} \sum_{i=1}^{N} (0 - 2\, e_2^T \alpha_i + 2 p_{i2} e_2^T e_2) \right) = 0$$

We know $e_2^T e_2 = 1$. Therefore, we have,

$$p_{i2} = e_2^T \alpha_i$$

Substituting for α_i,

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

Where $e_2^T e_1 = 0$, gives us

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

(b).

We are given,

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

Let's perform $\frac{\partial \tilde{J}}{\partial e_2} = 0$

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

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

Multiplying by e_1^T

$$e_1^T(-2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1) = 0$$

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

Since, $S = S^T$,

$$(-2Se_1^T e_2 + 2\lambda_2 e_1^T e_2 + \lambda_{12} e_1^T e_1) = 0$$

Now, $e_1^T e_2 = 0$, $e_2^T e_1 = 0$ and $e_1^T e_1 = 0$

$$(0 + 0 + \lambda_{12}) = 0 \; or \; \lambda_{12} = 0$$

Substituting for $\lambda_{12} = 0$ in $(-2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1) = 0$

$$(-2Se_2 + 2\lambda_2 e_2) = 0$$

Or,

$$Se_2 = \lambda_2 e_2$$

Hence, the value of e_2 minimising $\tilde{J}$ is given by second largest eigenvector of S

1.2 A Real Example

(a)

We are 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 by n-1 (= 99), we have

$$S = \begin{bmatrix} .9235 & 1.7366 & 3.01 \\ 1.7366 & 3.777 & 5.5072 \\ 3.01 & 5.5072 & 13.1036 \end{bmatrix}$$

Let us consider $\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}$. Then,

$$S - \lambda I = \begin{bmatrix} .9235 - \lambda_1 & 1.7366 & 3.01 \\ 1.7366 & 3.777 - \lambda_2 & 5.5072 \\ 3.01 & 5.5072 & 13.1036 - \lambda_3 \end{bmatrix}$$

Equating $\det(S - \lambda I) = 0$ and solving, we get,

$$\lambda_1 = 1626.5, \lambda_2 = 129, and\ \lambda_3 = 7.1$$

Or

$$\lambda = \begin{bmatrix} 1626.5 \\ 129 \\ 7.1 \end{bmatrix}$$

We can now compute the eigenvectors using,

$$(S - \lambda_i I)X_i = 0$$

where λ_i are the different eigenvalues and X_i are the corresponding eigenvectors.

Solving using the different eigenvalues, we get,

$$\mathrm{X}_1 = \begin{bmatrix} 0.218 \\ 0.414 \\ 0.884 \end{bmatrix}, \mathrm{X}_2 = \begin{bmatrix} 0.247 \\ 0.853 \\ -0.461 \end{bmatrix} and\ \mathrm{X}_3 = \begin{bmatrix} 0.944 \\ -0.318 \\ -0.084 \end{bmatrix}$$

(b)

Consider the contributions from the principal components of the different eigenvectors

We can compute them as $= \frac{\lambda_i}{\sum_i \lambda_i}$

For X_1, we have $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = \frac{1626.5}{1626.5+129+7.1} = 0.92$

For X_2, we have $\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \frac{129}{1626.5+129+7.1} = 0.07$

For X_3, we have $\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} = \frac{7.1}{1626.5+129+7.1} = 0.004$

As we can see, $\lambda_1$contributes the highest to the variance, followed by λ_2 and λ_3. The contribution from λ_1 and λ_2 are significant and so we can retain them. But the contribution from λ_3 is negligible and hence X_3 can be skipped.

(c)

The eigenvalue X_1 all have a positive value, which indicate that the birds with larger size tend to have larger length, wingspan and weight. It has max absolute value for weight, hence any change in this dimension will affect the size more than any other dimension.

Similarly, The eigenvalue X_2 has the highest values in the second and third entries (considering the absolute values). The values of second and third entries are opposite in sign which indicates that the birds with larger wingspans have smaller weights and vice-versa. It has max value for wingspan (absolute value)., hence any change in this dimension will affect the size more than any other dimension.

Lastly, the eigenvector X_3 has the max value for length (absolute value), which represents that any change in that dimension is going to affect the size more than other 2 dimensions.

2. Hidden Markov Model

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

Given, O = ACCGTA

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

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

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

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

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

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

Hence, we get

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

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

$$\alpha_2(1) = P(O_2|S_2 = 1) \times \sum_i a_{i1} \alpha_1(j) = b_{1C} \times \left(a_{11}\alpha_1(1) + a_{21}\alpha_1(2)\right)$$
$$= 0.2 \times (0.7 \times 0.24 + 0.4 \times 0.08) = 0.04$$

Computing others similarly, we get

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

$$\alpha_3(1) = b_{1C} \times \left(a_{11}\alpha_2(1) + a_{21}\alpha_2(2)\right) = 0.00944$$

$$\alpha_3(2) = b_{2C} \times \left(a_{12}\alpha_2(1) + a_{22}\alpha_2(2)\right) = 0.01632$$

$$\alpha_4(1) = b_{1G} \times \left(a_{11}\alpha_3(1) + a_{21}\alpha_3(2)\right) = 0.00394$$

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

$$\alpha_5(1) = b_{1T} \times \left(a_{11}\alpha_4(1) + a_{21}\alpha_4(2)\right) = 0.0003263$$

$$\alpha_5(2) = b_{2T} \times \left(a_{12}\alpha_4(1) + a_{22}\alpha_2(2)\right) = 0.0005819$$

$$\alpha_6(1) = b_{1A} \times \left(a_{11}\alpha_5(1) + a_{21}5(2)\right) = 0.0001844$$

$$\alpha_6(2) = b_{2A} \times \left(a_{12}\alpha_5(1) + a_{22}\alpha_5(2)\right) = 0.0000894$$

Therefore, probability of the observed sequence,

$$P(O;\ \theta) = \alpha_6(1) + \alpha_6(2) = 0.0001844 + 0.0000894 = 0.00002738$$

(b)

For a given state at a given time, the likelihood is given by:

$$\gamma_t(j) = P\left(X_t = S_j \middle| O_{1\ldots T}\right) = \frac{\alpha_t(j)\beta_t(j)}{\sum_i \alpha_t(i)\beta_t(i)}$$

Where

$$\beta_j = P(O_{t+1:T}|X_t = S_j)$$

Using the above and $\beta_6(1) = 1 \, and \, \beta_6(2) = 1$, we can compute

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

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

$$\beta_4(1) = \alpha_{11}b_{1A}\beta_5(1) + \alpha_{12}b_{2A}\beta_5(2) = 0.049$$

$$\beta_4(2) = \alpha_{21}b_{1A}\beta_5(1) + \alpha_{22}b_{2A}\beta_5(2) = 0.064$$

$$P(X_6 = S_1|O;\theta) = \frac{\alpha_6(1)\beta_6(1)}{\alpha_6(1)\beta_6(1) + \alpha_6(2)\beta_6(2)} = \frac{0.0001872}{0.0002772} = 0.67355$$

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

(c)

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

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

(d)

We have, from Viterbi algorithm,

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

We can compute the values as

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

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

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

$\delta_2(2) = max(\delta_1(1)b_{2C}a_{11}, b_{2C}\delta_1(2)a_{21}) = 0.0288$

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

$\delta_3(2) = max(\delta_2(1)b_{2C}a_{11}, b_{2C}\delta_2(2)a_{21}) = 0.006912$

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

$\delta_4(2) = max(\delta_3(1)b_{2G}a_{11}, b_{2G}\delta_3(2)a_{21}) = 0.0004147$

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

$\delta_5(2) = max(\delta_4(1)b_{2T}a_{11}, b_{2C}\delta_4(2)a_{21}) = 0.00008890$

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

$\delta_6(2) = max(\delta_5(1)b_{2A}a_{11}, b_{2A}\delta_5(2)a_{21}) = 0.00001066$

(e)

From earlier, we have

$$P(X_6 = S_1|O;\theta) = 0.67355$$

$$P(X_6 = S_2|O;\theta) = 0.32645$$

Now, let's compute the likelihood values for each possibility in {A, C, G, T},

For $O_7 = A$:

$$\begin{aligned} P(A|O;\theta) &= P(X_7 = S_1|\, O;\theta) + P(X_7 = S_2|\, O;\theta) \\ &= [b_{1A}P(X_6 = S_1|O;\theta)a_{11} + b_{1A}P(X_6 = S_2|O;\theta)a_{21}] \\ &\qquad + [b_{2A}P(X_6 = S_1|O;\theta)a_{12} + b_{2A}P(X_6 = S_2|O;\theta)a_{22}] \\ &= 0.3204 \end{aligned}$$

For $O_7 = C$:

$$\begin{aligned} P(C|O;\theta) &= P(X_7 = S_1|\, O;\theta) + P(X_7 = S_2|\, O;\theta) \\ &= [b_{1C}P(X_6 = S_1|O;\theta)a_{11} + b_{1C}P(X_6 = S_2|O;\theta)a_{21}] \\ &\qquad + [b_{2C}P(X_6 = S_1|O;\theta)a_{12} + b_{2C}P(X_6 = S_2|O;\theta)a_{22}] \\ &= 0.2795 \end{aligned}$$

For $O_7 = G$:

$$\begin{aligned} P(G|O;\Theta) &= P(X_7 = S_1|\, O;\theta) + P(X_7 = S_2|\, O;\theta) \\ &= [b_{1G}P(X_6 = S_1|O;\theta)a_{11} + b_{1G}P(X_6 = S_2|O;\theta)a_{21}] \\ &\qquad + [b_{2G}P(X_6 = S_1|O;\theta)a_{12} + b_{2G}P(X_6 = S_2|O;\theta)a_{22}] \\ &= 0.2204 \end{aligned}$$

For $O_7 = T$:

$$\begin{aligned} P(T|O;\theta) &= P(X_7 = S_1|\, O;\theta) + P(X_7 = S_2|\, O;\theta) \\ &= [b_{1T}P(X_6 = S_1|O;\theta)a_{11} + b_{1T}P(X_6 = S_2|O;\theta)a_{21}] \\ &\qquad + [b_{2T}P(X_6 = S_1|O;\theta)a_{12} + b_{2T}P(X_6 = S_2|O;\theta)a_{22}] \\ &= 0.1795 \end{aligned}$$

Therefore, the 7^{th} observation is most likely to be **A**

Collaboration

Collaborated on thoughts and ideas with **Adarsha Desai** and **Mahesh Pottippala Subrahmanya**