Assignment6

Arpita Agrawal

November 2016

1 Principal Component Analysis

1.1 Derivation of Second Principal Component

(a) $J = \frac{1}{N}\sum_{i=1}^{N}(x_i - p_{i1}e_1 - p_{i2}e_2)^T(x_i - p_{i1}e_1 - p_{i2}e_2), \qquad ||e_1||_2 = 1, ||e_2||_2 = 1, e_1^T e_2 = 0$

$$\frac{\partial J}{\partial p_{i2}} = \frac{2}{N}e_2^T(x_i - p_{i1}e_1 - p_{i2}e_2) = 0$$

Since, e_1, e_2 are orthogonal base vectors, $||e_2||_2 = 1, \quad e_1^T e_2 = 1$

$$\frac{2}{N}(e_2^T x_i - p_{i1}e_1^T e_2 - p_{i12}e_2^T e_2) = \frac{2}{N}(e_2^T x_i - 0 - p_{i2}) = 0$$

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

(b)

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

Value of e_2, that minimises the above cost function:

$$\frac{\partial J}{\partial e_2} = -Se_2 - S^T e_2 + 2\lambda_2 e_2 + \lambda_{12}e_1 = 0$$

Since S is a symmetric matrix, $S^T = S$,

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

Multiplying the equation with e_1^T:

$$e_1^T(-2Se_2 + 2\lambda_2 e_2 + \lambda_{12}e_1) = -2e_1^T S e_2 + 2e_1^T e_2\lambda_2 + e_1^T e_1\lambda_{12} = -0 + 0 + 1 * \lambda_{12} = 0$$

Substituting the value of $\lambda_{12} = 0$,

$$Se_2 = \lambda_2 e_2$$

1.2 A Real Example

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

$$S - \lambda I = 0, \qquad \begin{vmatrix} 91.43 & 171.92 & 297.99 \\ & 373.92 & 545.21 \\ & & 1297.26 \end{vmatrix} - \lambda \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 0$$

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

Taking determinant, Using MATLAB to compute the values:

$$\lambda_1 = 7.0975, \quad \lambda_2 = 128.9861, \quad \lambda_3 = 1626.5264$$

Substituting the above λ(eigen) values, we get the following eigenvectors:

$$v_1 = \begin{vmatrix} 0.25 \\ 0.85 \\ -0.46 \end{vmatrix}, \quad v_2 = \begin{vmatrix} 0.22 \\ 0.41 \\ 0.88 \end{vmatrix}, \quad v_3 = \begin{vmatrix} 0.94 \\ -0.32 \\ -0.08 \end{vmatrix}$$

b) Looking at the values of λ computed above, we see that λ_1 is much smaller than λ_1 and λ_3. The most variation is seen in λ_3, which shows that a lot of variation in the data has been incorporated in this new feature λ_3 which will be our first principal component. Next λ_2, also has a comparatively big value and can be taken as the Second Principal Component. The third $_3$, has a very small value and hence shows very less variation in the data. Hence, this orthonormal vector λ_3 can be omitted.

c) Interpretation of λ values that contains most of the information.

$$v_2 = \begin{vmatrix} 0.22 \\ 0.41 \\ 0.88 \end{vmatrix}$$

In the above vector all the three values are positive and define features - length, wingspan and weight. Since, all three of them are positive, we can say that as size of the bird increases i.e the length, wingspan and weight increase the size of the bird increases. So, this orthonormal vector can be seen as a feature defining size which generally increases when all the three features -(length, wingspan and weight) features increases. In the vector, weight has the highest value, that means the weight feature tends to impact the birds shape more and similarly, wingspan is the next feature that defines the shape of the bird more after weight, the more the wingspan the bigger the shape of the bird.

For the next principal component, there are 2 positive values and one negative value, in this vector wingspan and weight has highest absolute values, so these two parameters affect the data in this base vector. But, they are of opposite signs, so if weight increases then wingspan decreases, and if wingspan increases then weight decreases

2 Principal Component Analysis

2.1 $a_{11} = 0.7, a_{12} = 0.3; a_{21} = 0.4, a_{22} = 0.6.$
$b_{1A} = 0.4, b_{1C} = 0.2, b_{1G} = 0.2, b_{1T} = 0.1$ $b_{2A} = 0.2, b_{2C} = 0.4, b_{2G} = 0.1, b_{2T} = 0.3$

Observed Sequence O = ACCGTA

a) P(O; θ) = $\sum_{j=1}^{2} \alpha(j)$

$\alpha_1(1) = \pi_1 * b_{1A} = 0.6 * 0.4 = 0.24$
$\alpha_1(2) = \pi_2 * b_{2A} = 0.4 * 0.2 = 0.08$
$\alpha_2(1) = b_{1C} * (a_1(1)\alpha_1(1) + a_2(1)\alpha_1(2)) = 0.2 * (0.7 * 0.24 + 0.4 * 0.08) = 0.04$

$\alpha_2(2) = b_{2C} * (a_1(2)\alpha_1(1) + a_2(2)\alpha_1(2)) = 0.4 * (0.3 * 0.24 + 0.6 * 0.08) = 0.048$

$\alpha_3(1) = b_{1C} * (a_1(1)\alpha_2(1) + a_2(1)\alpha_2(2)) = 0.2 * (0.7 * 0.04 + 0.4 * 0.048) = 0.00944$
$\alpha_3(2) = b_{2C} * (a_1(2)\alpha_2(1) + a_2(2)\alpha_2(2)) = 0.4 * (0.3 * 0.04 + 0.6 * 0.048) = 0.01632$

$\alpha_4(1) = b_{1G} * (a_1(1)\alpha_3(1) + a_2(1)\alpha_3(2)) = 0.3 * (0.7 * 0.00944 + 0.4 * 0.01632) = 0.0039408$
$\alpha_4(2) = b_{2G} * (a_1(2)\alpha_3(1) + a_2(2)\alpha_3(2)) = 0.1 * (0.3 * 0.00944 + 0.6 * 0.01632) = 0.0012624$

$\alpha_5(1) = b_{1T} * (a_1(1)\alpha_4(1) + a_2(1)\alpha_4(2)) = 0.1 * (0.7 * 0.0039408 + 0.4 * 0.0012624) = 0.000326352$
$\alpha_5(2) = b_{2T} * (a_1(2)\alpha_4(1) + a_2(2)\alpha_4(2)) = 0.3 * (0.3 * 0.0039408 + 0.6 * 0.0012624) = 0.000581904$

$$\alpha_6(1) = b_{1A} * (a_1(1)\alpha_5(1) + a_2(1)\alpha_5(2)) = 0.4 * (0.7 * 0.000326352 + 0.4 * 0.000581904) = 0.000184483$$
$$\alpha_6(2) = b_{2A} * (a_1(2)\alpha_5(1) + a_2(2)\alpha_5(2)) = 0.2 * (0.3 * 0.000326352 + 0.6 * 0.000581904) = 8.94096E - 05$$

P(O;θ) $= \alpha_6(1) + \alpha_6(2) = 0.000184483 + 8.94096E - 05 = 0.000273893$

b) Calculate $P(X_6 = S_i|O;\theta) fori = 1, 2$
Using backward Algorithm, $\beta_{t-1} = \sum_{j=1}^{2} \beta_t a_{ij} P(b_t|X_t = S_j)$

$$\beta_6(1) = 1$$
$$\beta_6(2) = 1$$
$$\beta_5(1) = \beta_6(1)(a_{11}b_{1A} + a_{12}b_{2A}) = 1(0.7 * 0.4 + 0.3 * 0.2) = 0.34$$
$$\beta_5(2) = \beta_6(2)(a_{21}b_{1A} + a_{22}b_{2A}) = 1(0.4 * 0.4 + 0.6 * 0.2) = 0.28$$
$$\beta_4(1) = \beta_5(1)(a_{11}b_{1T} + a_{12}b_{2T}) = 0.34(0.7 * 0.1 + 0.3 * 0.3) = 0.049$$
$$\beta_4(2) = \beta_5(2)(a_{21}b_{1T} + a_{22}b_{2T}) = 0.28(0.4 * 0.1 + 0.6 * 0.3) = 0.064$$

$$P(X_6 = S_i|O,\theta) = \frac{\alpha_6(S_i)\beta_6(S_i)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)}$$

$$P(X_6 = S_1|O,\theta) = \frac{\alpha_6(S_1)\beta_6(S_1)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} = \frac{0.000184483}{0.000273893} = 0.673559875$$

$$P(X_6 = S_2|O,\theta) = \frac{\alpha_6(S_2)\beta_6(S_2)}{\alpha_6(S_1)\beta_6(S_1) + \alpha_6(S_2)\beta_6(S_2)} = \frac{8.94096E - 05}{0.000273893} = 0.326440125$$

c)

$$P(X_4 = S_i|O,\theta) = \frac{\alpha_4(S_i)\beta_4(S_i)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)}$$

$$P(X_4 = S_1|O,\theta) = \frac{\alpha_4(S_1)\beta_4(S_1)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} = \frac{0.000193099}{0.000273893} = 0.705017437$$

$$P(X_4 = S_2|O,\theta) = \frac{\alpha_4(S_2)\beta_4(S_2)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} = \frac{8.07936E - 05}{0.000273893} = 0.294982563$$

d) Findin the optimal path, Using Viterbi Algorithm

$$\delta_t(j) = max_i \delta_t(i) a_{ij} P(x_t|Z_t = s_i)$$
$$\delta_1(1) = \pi_1 b_{1A} = 0.6$$
$$\delta_1(2) = \pi_2 b_{2A} = 0.4$$

$$\delta_2(1) = b_{1C}(max(\delta_1(1)a_{11}, \delta_1(2)a_{21})) = 0.084$$
$$\delta_2(2) = b_{2C}(max(\delta_1(1)a_{12}, \delta_1(2)a_{22})) = 0.096$$

$$\delta_3(1) = b_{1C}(max(\delta_3(1)a_{11}, \delta_3(2)a_{21})) = 0.01176$$
$$\delta_3(2) = b_{2C}(max(\delta_3(1)a_{12}, \delta_3(2)a_{22})) = 0.02304$$

$$\delta_4(1) = b_{1C}(max(\delta_4(1)a_{11}, \delta_4(2)a_{21})) = 0.0027648$$
$$\delta_4(2) = b_{2C}(max(\delta_4(1)a_{12}, \delta_4(2)a_{22})) = 0.0013824$$

$$\delta_5(1) = b_{1C}(max(\delta_5(1)a_{11}, \delta_4(2)a_{21})) = 0.00193536$$
$$\delta_5(2) = b_{2C}(max(\delta_5(1)a_{12}, \delta_4(2)a_{22})) = 0.000248832$$

$$\delta_6(1) = b_{1C}(max(\delta_6(1)a_{11}, \delta_4(2)a_{21})) = 5.41901E - 05$$
$$\delta_6(2) = b_{2C}(max(\delta_6(1)a_{12}, \delta_4(2)a_{22})) = 2.98598E - 05$$

So, the steps followed in terms of hidden states will be :

S1 - S2 - S2 - S1 - S2 - S1

e) $P(O_7|O;\theta)$

$$P(O_7|O;\theta) = \sum_{i=1}^{2} P(O_7, X_7 = S_i|O)$$

$$P(O_7|O;\theta) = \sum_{i=1}^{2} P(O_7|X_7 = S_i) * \sum_{j=1}^{2} P(X_7 = S_i|X_6 = S_j)P(X_6 = S_j|O)$$

$$b_{1k} * (P(X_6 = S_1|\theta) * a_{11} + P(X_6 = S_2|\theta) * a_{21}) + b_{2k} * (P(X_6 = S_1|\theta) * a_{12} + P(X_6 = S_2|\theta) * a_{22})$$

$$= 0.602067962 * b_{1k} + 0.397932038 * b_{2k}$$

$$P(O_7 = A|\theta) = 0.602067962 * 0.4 + 0.397932038 * 0.2 = 0.320413592$$
$$P(O_7 = A|\theta) = 0.602067962 * 0.2 + 0.397932038 * 0.4 = 0.279586408$$
$$P(O_7 = A|\theta) = 0.602067962 * 0.3 + 0.397932038 * 0.1 = 0.220413592$$
$$P(O_7 = A|\theta) = 0.602067962 * 0.1 + 0.397932038 * 0.3 = 0.179586408$$

Therefore the next probable observed sequence will be A, having the max probability amongst the four possible code