

## CSCI 567 HW # 6

Mohmmad Suhail Ansari  
USC ID: 8518586692  
e-mail: mohmmada@usc.edu

November 21, 2016

**Sol. 1.1** Given

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

Taking the derivative w.r.t  $p_{i2}$

$$\frac{\partial J}{\partial p_{i2}} = \frac{1}{N} 2(-e_2^T)(x_i - p_{i1}e_1 - p_{i2}e_2) = 0$$

$$= -e_2^T x_i + p_{i1}e_2^T e_1 + p_{i2}e_2^T e_2 = 0$$

Since, we know that  $e_2^T e_1 = 0$  and  $\|e_2\|_2 = e_2^T e_2 = 1$ , we get

$$= p_{i2} - e_2^T x_i = 0$$

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

**Sol. 1.2** Given

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

Taking the derivative w.r.t  $e_2$ , we get

$$\frac{\partial \tilde{J}}{\partial e_2} = -(S + S^T)e_2 + 2\lambda_2 e_2 + \lambda_{12}e_1 = 0 \quad (1)$$

Since,  $S$  is symmetric, i.e.  $S = S^T$ , we get

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

multiplying with  $e_1^T$  from left, we get

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

since,  $e_1^T e_2 = 0$  and  $e_1^T e_1 = 1$

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

We also know, that  $(S e_1)^T e_2 = 0$  because  $e_1^T e_2 = 0$ , hence

$$\lambda_{12} = 0$$

and therefore from equation (1), we get

$$S e_2 = \lambda_2 e_2$$

and this proves that the value of  $e_2$  which minimizes  $\tilde{J}$  is given by the second largest eigen vector of  $S$ .

**Sol. 1.3** Using [numpy.linalg.eig](#), we get the following eigenvalues

$$\lambda_1 = 1626.5264, \quad \lambda_2 = 128.9860, \quad \lambda_3 = 7.0974$$

and the following eigenvectors

$$v_1 = \begin{bmatrix} 0.2179 \\ 0.4144 \\ 0.8835 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} -0.2466 \\ -0.8525 \\ 0.4607 \end{bmatrix}$$

$$v_3 = \begin{bmatrix} 0.9442 \\ -0.3183 \\ -0.0835 \end{bmatrix}$$

**Sol. 1.4** Yes, we can omit the direction of the vector  $v_3$ , since  $\lambda_1$  and  $\lambda_2$  account for most the information. It can be shown that for  $\lambda_1$ ,  $\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = 0.9227$ , i.e.  $\lambda_1$  is account for 92.28% of the information and for  $\lambda_2$ ,  $\frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = 0.0731$ , i.e.  $\lambda_2$  is account for 7.32% of the information, which makes  $\lambda_3$  account for only 0.4% for the information, and hence can be omitted.

**Sol. 1.5** For  $v_1$ , since all the elements of the vector have the same sign, it suggest that all the attributes increase and decrease w.r.t to each other, i.e. for example if the bird has large length, then it will also have large wing span and a higher weight and vice versa. This makes sense since, a large bird will tend to have all three attributes larger than a small bird.

**Sol. 2.1** To calculate  $P(O; \theta)$ , we first calculate the following

$$\alpha_1(1) = \pi_1 \times b_{1A} = 0.6 \times 0.4 = 0.24$$

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

$$\alpha_2(1) = b_{1C}(\alpha_1(1)a_{11} + \alpha_1(2)a_{21}) = 0.2 \times (0.24 \times 0.7 + 0.08 \times 0.4) = 0.04$$

$$\alpha_2(2) = b_{2C}(\alpha_1(1)a_{12} + \alpha_1(2)a_{22}) = 0.4 \times (0.24 \times 0.3 + 0.08 \times 0.6) = 0.048$$

$$\alpha_3(1) = b_{1C}(\alpha_2(1)a_{11} + \alpha_2(2)a_{21}) = 0.2 \times (0.04 \times 0.7 + 0.048 \times 0.4) = 0.00944$$

$$\alpha_3(2) = b_{2C}(\alpha_2(1)a_{12} + \alpha_2(2)a_{22}) = 0.4 \times (0.04 \times 0.3 + 0.048 \times 0.6) = 0.01632$$

$$\alpha_4(1) = b_{1G}(\alpha_3(1)a_{11} + \alpha_3(2)a_{21}) = 0.3 \times (0.00944 \times 0.7 + 0.01632 \times 0.4) = 0.0039408$$

$$\alpha_4(2) = b_{2G}(\alpha_3(1)a_{12} + \alpha_3(2)a_{22}) = 0.1 \times (0.00944 \times 0.3 + 0.01632 \times 0.6) = 0.0012624$$

$$\alpha_5(1) = b_{1T}(\alpha_4(1)a_{11} + \alpha_4(2)a_{21}) = 0.1 \times (0.0039408 \times 0.7 + 0.0012624 \times 0.4) = 0.000326352$$

$$\alpha_4(2) = b_{2T}(\alpha_4(1)a_{12} + \alpha_4(2)a_{22}) = 0.3 \times (0.0039408 \times 0.3 + 0.0012624 \times 0.6) = 0.000581904$$

$$\alpha_6(1) = b_{1A}(\alpha_5(1)a_{11} + \alpha_5(2)a_{21}) = 0.4 \times (0.000326352 \times 0.7 + 0.000581904 \times 0.4) = 0.0001844832$$

$$\alpha_6(2) = b_{2A}(\alpha_5(1)a_{12} + \alpha_5(2)a_{22}) = 0.2 \times (0.000326352 \times 0.3 + 0.000581904 \times 0.6) = 0.0000894096$$

Finally,

$$P(O; \theta) = \alpha_6(1) + \alpha_6(2) = 0.0001844832 + 0.0000894096 = 0.0002738928$$

**Sol. 2.2** To calculate  $P(X_6 = S_i | O; \theta)$ , we first calculate the following

$$\beta_6(1) = 1$$

$$\beta_6(2) = 1$$

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

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

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

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

$$\beta_3(1) = (b_{1G}a_{11}\beta_4(1) + b_{2G}a_{12}\beta_4(2)) = 0.01221$$

$$\beta_3(2) = (b_{1G}a_{21}\beta_4(1) + b_{2G}a_{22}\beta_4(2)) = 0.00972$$

$$\beta_2(1) = (b_{1C}a_{11}\beta_3(1) + b_{2C}a_{12}\beta_3(2)) = 0.0028758$$

$$\beta_2(2) = (b_{1C}a_{21}\beta_3(1) + b_{2C}a_{22}\beta_3(2)) = 0.0033096$$

$$\beta_1(1) = (b_{1C}a_{11}\beta_2(1) + b_{2C}a_{12}\beta_2(2)) = 0.000799764$$

$$\beta_1(2) = (b_{1C}a_{21}\beta_2(1) + b_{2C}a_{22}\beta_2(2)) = 0.001024368$$

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

$$= \frac{0.0001844832 \times 1}{0.0001844832 \times 1 + 0.0000894096 \times 1} = 0.67355987452$$

$$P(X_6 = S_2|O; \theta) = 1 - P(X_6 = S_1|O; \theta) = 1 - 0.67355987452 = 0.32644012548$$

**Sol. 2.3**

$$P(X_4 = S_1|O; \theta) = \frac{\alpha_4(S_1)\beta_4(S_1)}{\alpha_4(S_1)\beta_4(S_1) + \alpha_4(S_2)\beta_4(S_2)} = \frac{0.0039408 \times 0.049}{0.0039408 \times 0.049 + 0.0012624 \times 0.064} = 0.70501743747$$

$$P(X_4 = S_2|O; \theta) = 1 - P(X_4 = S_1|O; \theta) = 1 - 0.70501743747 = 0.29498256253$$

**Sol. 2.4**

$$P(X_1 = S_1|O; \theta) = \frac{0.24 \times 0.000799764}{0.24 \times 0.000799764 + 0.08 \times 0.001024368} = 0.70079739226$$

$$P(X_1 = S_2|O; \theta) = 0.29920260773$$

$$P(X_2 = S_1|O; \theta) = \frac{0.04 \times 0.0028758}{0.04 \times 0.0028758 + 0.048 \times 0.0033096} = 0.41998913443$$

$$P(X_2 = S_2|O; \theta) = 0.58001086556$$

$$P(X_3 = S_1|O; \theta) = \frac{0.00944 \times 0.01221}{0.00944 \times 0.01221 + 0.01632 \times 0.00972} = 0.42083033946$$

$$P(X_3 = S_2|O; \theta) = 0.57916966053$$

$$P(X_4 = S_1|O; \theta) = 0.70501743747$$

$$P(X_4 = S_2|O; \theta) = 0.29498256253$$

$$P(X_5 = S_1|O; \theta) = \frac{0.000326352 \times 0.34}{0.000326352 \times 0.34 + 0.000581904 \times 0.28} = 0.40512083559$$

$$P(X_5 = S_2|O; \theta) = 0.5948791644$$

$$P(X_6 = S_1|O; \theta) = 0.67355987452$$

$$P(X_6 = S_2|O; \theta) = 0.32644012548$$

So, the most likely sequence is

$$S_1 S_2 S_2 S_1 S_2 S_1$$

**Sol. 2.4**

$$\begin{aligned}P(O_7 = A|O; \theta) &= P(X_6 = S_1|O; \theta) \times b_{1A} + P(X_6 = S_2|O; \theta) \times b_{2A} \\&= 0.67355987452 \times 0.4 + 0.32644012548 \times 0.2 = 0.3347119749\end{aligned}$$

$$\begin{aligned}P(O_7 = T|O; \theta) &= P(X_6 = S_1|O; \theta) \times b_{1T} + P(X_6 = S_2|O; \theta) \times b_{2T} \\&= 0.67355987452 \times 0.1 + 0.32644012548 \times 0.3 = 0.16528802509\end{aligned}$$

$$\begin{aligned}P(O_7 = G|O; \theta) &= P(X_6 = S_1|O; \theta) \times b_{1G} + P(X_6 = S_2|O; \theta) \times b_{2G} \\&= 0.67355987452 \times 0.3 + 0.32644012548 \times 0.1 = 0.2347119749\end{aligned}$$

$$\begin{aligned}P(O_7 = C|O; \theta) &= P(X_6 = S_1|O; \theta) \times b_{1C} + P(X_6 = S_2|O; \theta) \times b_{2C} \\&= 0.67355987452 \times 0.2 + 0.32644012548 \times 0.4 = 0.26528802509\end{aligned}$$

$$O_7 = \operatorname{argmax}_O P(O|O; \theta) = A$$