MACHINE LEARNING
ASSIGNMENT 6

AJAY KUMAR LOGANATHAN RAVICHANDRAN
USC ID: 1669 4689 06

1)

Principal Component Analysis

Given: N zero mean data points $x_i \in R^{D+1}$ and $S = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T \in R^{D \times D}$

1.1)

Derivation of Second Principal Component

a)

Given:

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 orthogonal vector basis for dimensionality reduction

$\Rightarrow \ ||e_1||_2 = 1, \ ||e_2||_2 = 1 \ , \ e_1^T e_2 = 0$

Coefficients $\Rightarrow \ p_{i1}, p_{i2}$

To Find:

$\frac{\partial J}{\partial p_{i2}} = 0 \ yields \ p_{i2} = e_2^T x_i$

Solution:

J can be written in vector form as :

$J \quad = \frac{1}{N}\sum_{i=1}^{N} || \, x_i - p_{i1}e_1 - p_{i2}e_2||^2$

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

$\Rightarrow \ \frac{-2}{N}\sum_{i=1}^{N}(e_2^T x_i - p_{i1}e_1e_2^T - p_{i2}e_2e_2^T) = 0$

Since $||e_2||_2 = 1 \ , \ e_2^T e_2 = 0$

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

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

Where e_2 is an orthogonal vector

b)

Given:

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

λ_2 is the lagrange multiplier for equality constraint $e_2^T e_2 = 1$

λ_{12} is the lagrange multiplier for equality constraint $e_2^T e_1 = 0$

To find:

Show that the value of e_2 that minimizes the cost function J is given by the eigenvector associated with the second largest eigenvalue of S.

Solution:

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

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

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

Also since S being a symmetric matrix,

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

In order to find the second largest eigenvalue of S, when e_2 minimizes the cost function, we need to find the value of λ_{12}

Multiplying the above equation with e_1^T, we get,

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

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

As $(e_1 e_2^T) = 0$, $\|e_1\|_2 = 1$ *and* $e_1 e_1^T = 1$, *the above equation becomes*,

$-2S(0) + 2\lambda_2(0) + \lambda_{12}(1) = 0$

$\lambda_{12} = 0$

Substituting value of λ_{12} in $-2Se_2 + 2\lambda_2 e_2 + \lambda_{12} e_1 = 0$, we get,

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

$\Rightarrow Se_2 = \lambda_2 e_2$

Thus, e_2 is the second largest eigenvalue of S.

1.2)

A Real Example

Given:

Sample size (n) = 100 bird species

Factors (m) = 3 (Length (inches), wingspan (inches) , weight(ounces))

Covariance matrix (S) =

[91.43 171.92 297.99

373.92 545.21

1297.26]

a)

To find: Eigenvalues and orthonormal eigenvectors

Eigenvalues:

In order to get the eigenvalues, we solve the below equation,

$|S-\lambda I| = 0$

$$det\left(\begin{bmatrix} 91.43-\lambda & 171.92 & 297.99 \\ 171.92 & 373.92-\lambda & 545.21 \\ 297.99 & 545.21 & 1297.26-\lambda \end{bmatrix}\right) = \mathbf{0}$$

Calculating the determinant we get,

$(91.43-\lambda)[\lambda^2-1671.18\lambda + 187817.52] - [10411098.69-29556.48\lambda] + [88798.04\lambda - 5272015.24] = 0$

$91.43\lambda^2 - 152779.53\lambda + 17172155.85 - \lambda^3 + 1671.18\lambda^2 - 187817.52\lambda - 10411098.69 + 29556.48\lambda + 88798.04\lambda - 5272015.24 = 0$

$\Rightarrow \lambda^3 - 1762.61\lambda^2 + 222242.53\lambda - 1489041.92 = 0$

Solving for λ, *we get*

$\lambda_1 = 1626.537474034728$

$\lambda_2 = 128.97447554027394$

$\lambda_3 = 7.09805042499772$

Orthonormal eigenvector calculation:

In order to calculate the eigenvector (v_k), we need to solve the below equation:

$(S-\lambda_k I)v_k = 0$

Let v_1 ($[x\ y\ z]$) be the eigenvector associated with λ_1, then

$(S-\lambda_1 I)v_1 = 0$

Since $\lambda_1 = 1626.526424060077$,

$$\begin{bmatrix} 91.43-1626.526 & 171.92 & 297.99 \\ 171.92 & 373.92-1626.526 & 545.21 \\ 297.99 & 545.21 & 1297.26-1626.526 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}$$

Solving for $[x\ y\ z]$, *we get* $v_1 = [0.22 \quad 0.41 \quad 0.88]$

Let v_2 ($[x\ y\ z]$) be the eigenvector associated with λ_2, then

$(S-\lambda_2 I)v_2 = 0$

Since $\lambda_2 = 128.97447554027394$

$$\begin{bmatrix} 91.43-128.97 & 171.92 & 297.99 \\ 171.92 & 373.92-128.97 & 545.21 \\ 297.99 & 545.21 & 1297.26-128.97 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}$$

Solving for $[x\,y\,z]$, *we get* $v_2 = [0.25 \quad 0.85 \quad -0.46]$

Let $v_3([x\,y\,z])$ be the eigenvector associated with λ_3, then

$(S - \lambda_3 I)v_3 = 0$

Since $\lambda_3 = \ 7.09805042499772$

$$\begin{bmatrix} 91.43 - 7.09 & 171.92 & 297.99 \\ 171.92 & 373.92 - 7.09 & 545.21 \\ 297.99 & 545.21 & 1297.26 - 7.09 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}$$

Solving for $[x\,y\,z]$, *we get* $v_3 = [0.94 \ -0.32 \ -0.08]$

b)

To Find:

Is there any of the orthonormal directions that can be omitted without losing lot of information? If yes which one(s) and why?

Solution:

It can be observed from the calculated eigenvalues and eigenvectors that v_3 accounts for the least variance in the data as λ_3 is the least of the all the eigenvalues computed.

Percentage of variance accounted by each eigenvectors:

v_1 :

$$\frac{1626.537474034728}{1626.537474034728 + 128.97447554027394 + 7.09805042499772} \times 100 = 92.28\ \%$$

v_2 :

$$\frac{128.97447554027394}{1626.537474034728 + 128.97447554027394 + 7.09805042499772} \times 100 = 7.32\%$$

v_3 :

$$\frac{7.09805042499772}{1626.537474034728 + 128.97447554027394 + 7.09805042499772} \times 100 = 0.40\%$$

Since v_3 accounts for $0.40\,\%$ of variance in the data, loss of information by omitting the vector v_3 will be negligible.

Thus we can omit the orthonormal vector v_3.

c)

To find:

Interpret the eigenvector(s) that contain(s) the most of the information regarding this data.

Solution:

Out of the three eigenvectors, the eigenvectors that contain most of the information are

$v_1(\textit{accounts for } 92.28\% \textit{ of data}) \textit{ and } v_2(\textit{accounts for } 7.32\% \textit{ of data})$, where:

$v_1 = [0.22 \quad 0.41 \quad 0.88]$

$v_2 = [0.25 \quad 0.85 \quad -0.46]$

The three factors which denote the size of the bird in the data are : length, wingspan and weight.

- From v_1, we can interpret that since all three values have same sign, it could imply that the bird with larger size might possibly have large length, wingspan and weight.
- In v_1, the weight entry is highest among other entries. Thus weight is a significant factor contributing to the size of the bird implying that even a small change in weight will affect the size of the bird more than the corresponding change in length or wingspan.
- From v_2, we can observe that the absolute values of the wingspan and weight entries in v_2 is greatest. Since, they have opposite signs, they imply that birds with smaller wingspan will have larger weight and vice versa.

2)

Hidden Markov Model

Given:

Parameters $\Theta = \{\pi_i, a_{ij}, b_{ik}\}\ for\ i,j\ = 1,2\ and\ k \in \{A, C, G, T\}$

Initial state distribution π :

$\pi_1 = P(X_1 = S_1) = 0.6$

$\pi_2 = P(X_1 = S_2) = 0.4$

Transition Probabilities $a_{ij} = P(X_{t+1} = S_j | X_t = S_i)\ for\ any\ t \in N^+$

$a_{11} = 0.7,\ a_{12} = 0.3;\ a_{21} = 0.4,\ a_{22} = 0.6$

Emission Probabilities $b_{ik} = P(O_t = k | X_t = S_i)\ for\ any\ t \in N^+$

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

Given Sequence: $O = ACCGTA$

a)

To find:

Probability of an observed sequence $(P(O;\ \Theta))$

Solution:

Forward algorithm can be used to find $P(O;\ \Theta)$

Steps:

Hidden states: $S_1,\ S_2$

Initialization: $\alpha_1(j) = p(O_1 | S_1 = j)\, p(S_1 = j)$

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

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

Recursion:

Iterations: $t = 2, 3.....T$

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

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

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

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

$$\alpha_3(2) = b_{2C}(a_{12}\alpha_2(1)+a_{22}\alpha_2(2)) = 0.4(0.3*0.04+0.6*0.048) = 0.01632$$
$$\alpha_4(1) = b_{1G}(a_{11}\alpha_3(1)+a_{21}\alpha_3(2)) = 0.3(0.7*0.00944+0.4*0.01632) = 0.0039408$$
$$\alpha_4(2) = b_{2G}(a_{12}\alpha_3(1)+a_{22}\alpha_3(2)) = 0.1(0.3*0.00944+0.6*0.01632) = 0.0012624$$
$$\alpha_5(1) = b_{1T}(a_{11}\alpha_4(1)+a_{21}\alpha_4(2)) = 0.1(0.7*0.0039408+0.4*.0012624) = 0.000326352$$
$$\alpha_5(2) = b_{2T}(a_{12}\alpha_4(1)+a_{22}\alpha_4(2)) = 0.3(0.3*0.0039408+0.6*.0012624) = 0.000581904$$
$$\alpha_6(1) = b_{1A}(a_{11}\alpha_5(1)+a_{21}\alpha_5(2)) = 0.4(0.7*0.000326352+0.4*0.000581904) = 0.00018448$$
$$\alpha_6(2) = b_{2A}(a_{12}\alpha_5(1)+a_{22}\alpha_5(2)) = 0.2(0.3*0.000326352+0.6*0.000581904) = 0.00008940$$

Termination step:

$$p(\{O_t\}_{t=1..6}) = \sum_{j=1}^{2}\alpha_6(j)$$

$\alpha_6(1)+\alpha_6(2) = 0.0002738928$

b)

Filtering

To find:

$P(X_6 = S_i\,|O;\,\Theta)\; for\; i = 1,\,2$

Solution:

Likelihood of a state at a given time (t) is given by $\gamma_t(j)$, where

$$\gamma_t(j) = P(X_t = s_j\,|O_{1:T}) = \frac{\alpha_t(j)\beta_t(j)}{\sum_k \alpha_t(k)\beta_t(k)}$$

According to Backward algorithm,

$$\beta_t(j) = P(O_{t+1:T}\,|X_t = s_j)$$

$$\beta_{t-1}(i) = \sum_j \beta_t(j)a_{ij}p(O_t\,|X_t = S_j)$$

Where $\beta_T(j)$ for the base case is equal to 1.

Thus, $\beta_6(1) = \beta_6(2) = 1$

$$P(X_6 = S_1\,|O;\,\Theta) = \frac{\alpha_6(1)\beta_6(1)}{\sum_{j=1}^{2}\alpha_6(j)\beta_6(j)}$$
$$= \frac{0.0001844832\times 1}{(0.0001844832\times 1)+(0.0000894096\times 1)}$$
$$= 0.673559897$$

$$P(X_6 = S_2\,|O;\,\Theta) = \frac{\alpha_6(2)\beta_6(2)}{\sum_{j=1}^{2}\alpha_6(j)\beta_6(j)}$$
$$= \frac{0.0000894096\times 1}{(0.0001844832\times 1)+(0.0000894096\times 1)}$$
$$= 0.32644012$$

c)

Smoothing

To Find:

$P(X_4 = S_i | O; \Theta)$ *for* $i = 1, 2$

Solution:

We know,

$$P(X_4 = S_i | O; \Theta) \quad = \frac{\alpha_4(i)\beta_4(i)}{\sum_{j=1}^{2} \alpha_4(j)\beta_4(j)}$$

Where :

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

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

$$\begin{aligned}\beta_5(1) &= \beta_6(1_1)a_{11}b_{1A} + \beta_6(2)a_{12}b_{2A} \\ &= 1 \times 0.7 \times 0.4 + 1 \times 0.3 \times 0.2 \\ &= 0.34\end{aligned}$$

$$\begin{aligned}\beta_5(2) &= \beta_6(1)a_{21}b_{1A} + \beta_6(2)a_{22}b_{2A} \\ &= 1 \times 0.4 \times 0.4 + 1 \times 0.6 \times 0.2 \\ &= 0.28\end{aligned}$$

$$\begin{aligned}\beta_4(1) &= \beta_5(1)a_{11}b_{1T} + \beta_5(2)a_{12}b_{2T} \\ &= 0.34 \times 0.7 \times 0.1 + 0.34 \times 0.3 \times 0.3 \\ &= 0.049\end{aligned}$$

$$\begin{aligned}\beta_4(2) &= \beta_5(1)a_{21}b_{1T} + \beta_5(2)a_{22}b_{2T} \\ &= 0.34 \times 0.4 \times 0.1 + 0.28 \times 0.6 \times 0.3 \\ &= 0.064\end{aligned}$$

$$\begin{aligned}P(X_4 = S_1 | O; \Theta) \quad &= \frac{\alpha_4(1)\beta_4(2)}{\sum_{j=1}^{2} \alpha_4(j)\beta_4(j)} \\ &= \frac{0.0039408 * 0.049}{(0.0039408 * 0.049) + (0.0012624 * 0.064)} \\ &= 0.7050\end{aligned}$$

$$\begin{aligned}P(X_4 = S_2 | O; \Theta) \quad &= \frac{\alpha_4(2)\beta_4(2)}{\sum_{j=1}^{2} \alpha_4(j)\beta_4(j)} \\ &= \frac{0.0012624 * 0.064}{(0.0039408 * 0.049) + (0.0012624 * 0.064)} \\ &= 0.2950\end{aligned}$$

d)

More likely explanation

To Find:

$X = argmax_X P(X|O; \Theta)$

Solution:

According to Viterbi algorithm,

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

P(Most likely path) = $argmax_j\delta_{1...6}(j)$

Thus

$\delta_1(1) = \pi_1 b_{1A} = 0.6 \times 0.4 \ = \ 0.24$

$\delta_1(2) = \pi_2 b_{2A} = 0.4 \times 0.2 \ = \ 0.08$

$\delta_2(1) \ = \ max(b_{1C}a_{11}\delta_1(1), b_{1C}a_{21}\delta_1(2)) = max(0.0336, 0.0064) \ = \ 0.0336$

$\delta_2(2) \ = \ max(b_{2C}a_{12}\delta_1(1), b_{2C}a_{22}\delta_1(2)) \ = \ max(0.0288, 0.0192) = 0.0288$

$\delta_3(1) = max(b_{1C}a_{11}\delta_2(1), b_{1C}a_{21}\delta_2(2)) = max(\ 0.004704, 0.002304) = 0.004704$

$\delta_3(2) = \ max(b_{2C}a_{12}\delta_2(1), b_{2C}a_{22}\delta_2(2)) \ = \ max(0.006912, 0.004032) = 0.006912$

$\delta_4(1) = max(b_{1G}a_{11}\delta_3(1), b_{1G}a_{21}\delta_3(2)) = max(\ 0.00098784, 0.00083) = 0.00098784$

$\delta_4(2) = \ max(b_{2G}a_{12}\delta_3(1), b_{2G}a_{22}\delta_3(2)) \ = \ max(0.00041472, 0.00014) = 0.000414$

$\delta_5(1) = max(b_{1T}a_{11}\delta_4(1), b_{1T}a_{21}\delta_4(2)) = \ max(0.0000693, 0.0000164) = 0.0000693$

$\delta_5(2) = \ max(b_{2T}a_{12}\delta_4(1), b_{2T}a_{22}\delta_4(2)) \ = \ max(0.0000891, 0.0000738) = 0.0000891$

$\delta_6(1) = max(b_{1A}a_{11}\delta_5(1), b_{1A}a_{21}\delta_5(2)) = \ max(0.00001936, 0.0000142) = 0.00001936$

$\delta_6(2) = \ max(b_{2A}a_{12}\delta_5(1), b_{2A}a_{22}\delta_5(2)) \ = max(\ 0.000010692, 0.00000415) = 0.000010692$

Thus the Most likely sequence is:

$S_1, S_1, S_1, S_1, S_1, S_1$

e)

Prediction

To Find:

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,

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

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

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

$= 0.60206 \times 0.4 + 0.39793 \times 0.2$

$= 0.32041$

$P(O|O=C;\Theta) \quad = m_1 \times b_{1C} + m_2 \times b_{2C}$

$= 0.60206 \times 0.2 + 0.39793 \times 0.4$

$= 0.279584$

$P(O|O=G;\Theta) \quad = m_1 \times b_{1G} + m_2 \times b_{2G}$

$= 0.60206 \times 0.3 + 0.39793 \times 0.1$

$= 0.220411$

$P(O|O=T;\Theta) \quad = m_1 \times b_{1T} + m_2 \times b_{2T}$

$= 0.60206 \times 0.1 + 0.39793 \times 0.3$

$= 0.179585$

$P(O|O=A;\Theta)$ has the maximum probability

Most likely observation at state 7 would be **A**

COLLABORATORS:

SINDHUJHA SETHURAMAN

SHIV RANJINI RAJAGOPAL