USC CSCI 567 HOMEWORK 6 SOLUTIONS

ThammeGowda Narayanaswamy
EMail: tnarayan@usc.edu USCID: 2074-6694-39

November 21, 2016

1 Principal Component Analysis

1.1 Derivation of Second Principal Component

1.1.a

Given : The cost function

$$\begin{aligned}
J =& \frac{1}{N}\sum_{i=1}^{N}(x_i - p_{i1}e_1 - p_{i2}e_2)^T(x_i - p_{i1}e_1 - p_{i2}e_2) \\
=& \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - 2p_{i1}e_1^T x_i - 2p_{i2}e_2^T x_i + p_{i1}^2 e_1^T e_1 + 2p_{i1}p_{i2}e_1^T e_2 + p_{i2}^2 e_2^T e_2 \\
=& \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - 2p_{i1}e_1^T x_i - 2p_{i2}e_2^T x_i + p_{i1}^2 + p_{i2}^2 \qquad \because e_1^T e_1 = 1, e_2^T e_2 = 1, e_1^T e_2 = 0
\end{aligned} \tag{1}$$

By differentiation 1 with respect to p_{i2}, we get

$$\begin{aligned}
\frac{\partial J}{\partial p_{i2}} =& 0 \\
\implies \frac{\partial}{\partial p_{i2}}\frac{1}{N}\sum_{i=1}^{N} x_i^T x_i - 2p_{i1}e_1^T x_i - 2p_{i2}e_2^T x_i + p_{i1}^2 + p_{i2}^2 =& 0 \\
\frac{1}{N}\frac{\partial}{\partial p_{i2}}[x_i^T x_i - 2p_{i1}e_1^T x_i - 2p_{i2}e_2^T x_i + p_{i1}^2 + p_{i2}^2] =& 0 \qquad \because \text{other terms in sum are constant w.r.t a } p_{i2} \\
-2e_2^T x_i + 2p_{i2} =& 0 \\
\implies p_{i2} =& e_2^T x_i
\end{aligned}$$

1.1.b

$$\begin{aligned}
\tilde{J} = -e_2^T S e_2 + \lambda_2(e_2^T e_2 - 1) + \lambda_{12}(e_2^T e_1 - 0) \\
\implies \text{First order derivative w.r.t } e_2 \\
\frac{\partial \tilde{J}}{\partial e_2} = -2Se_2 + 2\lambda_2 e_2 + \lambda_{12}e_1 =& 0 \\
(S - \lambda_2 I)e_2 =& \frac{1}{2}\lambda_{12}e_1 \\
e_2 =& \frac{1}{2}(S - \lambda_2 I)^{-1}\lambda_{12}e_1
\end{aligned}$$

1.2 A Real Example

Given :
$m = 3$
$n = 100$

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

1.2.a

By factorizing the matrix S using singular value decomposition (SVD) method, we get

$$S = U\Sigma V^T = \begin{bmatrix} -0.218 & -0.247 & -0.944 \\ -0.414 & -0.853 & 0.318 \\ -0.884 & 0.461 & 0.084 \end{bmatrix} \times \begin{bmatrix} 1626.526 & 0 & 0 \\ 0. & 128.986 & 0 \\ 0. & 0 & 7.097 \end{bmatrix} \times \begin{bmatrix} -0.218 & -0.414 & -0.884 \\ -0.247 & -0.853 & 0.461 \\ -0.944 & 0.318 & 0.084 \end{bmatrix}$$

Where U is an orthonormal matrix formed by eigen vectors, Σ is a diagonal matrix having eigen values, and $V^T = U^T = U^{-1}$ for a symmetric matrix S.

Thus, the eigen values are: $\lambda_1 = 1626.526, \lambda_2 = 128.986, \lambda_3 = 7.097$

Eigen vectors (corresponding to eigen values in the same order) are

$$\begin{bmatrix} -0.218 \\ -0.414 \\ -0.884 \end{bmatrix}, \begin{bmatrix} -0.247 \\ -0.853 \\ 0.461 \end{bmatrix}, \begin{bmatrix} -0.944 \\ 0.318 \\ 0.084 \end{bmatrix}$$

- Note: The calculations of SVD method has been practiced as per the lecture from MIT 18.06SC Linear Algebra, Fall 2011 https://www.youtube.com/watch?v=cOUTpqlX-Xs .
- A computer program was used to perform the above calculations http://www.bluebit.gr/matrix-calculator/

1.2.b

Yes, the eigen vector corresponding to eigen value $\lambda_3 = 7.097$, i.e. $\begin{bmatrix} -0.944 \\ 0.318 \\ 0.084 \end{bmatrix}$ *may be* omitted. Reason: The eigenvalue λ_3 is much smaller compared to the other eigen values. Since the eign value represents the variation along the component (the associated eigen vector), smaller eign value $\implies$ the less variance $\implies$ less information captured by that component.

1.2.c

My interpretation for the eigenvectors that captures most information are as follows:

- Eigen vectors forms the core of principal component analysis. All the eign vectors are sorted based on the descending order of eigen values. Then the top k eigenvectors are used for projecting the data into newer space which has lesser number of dimensions
- Eigen values indicates the variance of data point when they are projected along the associated eigen vector. Higher variance is always preferred.
- The higher variance in the data set helps in easier statistical analysis, since (1) we can distinguish data points clearly when they are spread across, eg: classification. (2) we can reconstruct the original dimensions without loosing much information

2 Hidden Markov Model

Given:

The initial state probabilities: $\pi_1 = P(X_1 = s_1) = 0.6$ $\pi_2 = P(X_1 = s_2) = 0.4$

Transition probabilities :

	s_1	s_2
s_1	0.7	0.3
s_1	0.4	0.6

Observation probabilities:

	A	C	G	T
s_1	0.4	0.2	0.3	0.1
s_1	0.2	0.4	0.1	0.3

2.a Probability of Observed sequence

t	O_t	$\alpha_t(s_1)$	$\alpha_t(s_2)$
1	A	$\alpha_1(1) = P(A\|X_1 = s_1) \times P(X_1 = s_1)$ $= 0.6 \times 0.4$ $= 0.24$	$\alpha_1(2) = P(A\|X_1 = s_2) \times P(X_1 = s_2)$ $= 0.4 \times 0.2$ $= 0.08$
2	C	$\alpha_2(1) = P(C\|X_2 = s_1)[a_{11}\alpha_1(1) + a_{21}\alpha_1(2)]$ $= 0.2[0.7 \times 0.24 + 0.4 \times 0.08]$ $= 0.04$	$\alpha_2(2) = P(C\|X_2 = s_1)[a_{12}\alpha_1(1) + a_{22}\alpha_1(2)]$ $= 0.4[0.3 \times 0.24 + 0.6 \times 0.08]$ $= 0.048$
3	C	$\alpha_3(1) = 0.2[0.7 \times 0.04 + 0.4 \times 0.048]$ $= 0.00944$	$\alpha_3(2) = 0.4[0.3 \times 0.04 + 0.6 \times 0.048]$ 0.01632
4	G	$\alpha_4(1) = 0.3[0.7 \times 0.00944 + 0.4 \times 0.01632]$ $= 0.0039408$	$\alpha_4(2) = 0.1[0.3 \times 0.00944 + 0.6 \times 0.01632]$ $= 0.0012624$
5	T	$\alpha_5(1) = 0.1[0.7 \times 0.0039408 + 0.4 \times 0.0012624]$ $= 0.000326352$	$\alpha_5(2) = 0.3[0.3 \times 0.0039408 + 0.6 \times 0.0012624]$ $= 0.000581904$
6	A	$\alpha_6(1) =$ $0.4[0.7 \times 0.00032635 + 0.4 \times 0.0005819]$ $= 0.000184482$	$\alpha_6(2) =$ $0.2[0.3 \times 0.00032635 + 0.6 \times 0.0005819]$ $= 0.000089409$

$$P(O = ACCGTA|\Theta) = \alpha_6(1) + \alpha_6(2) = 0.000184482 + 0.000089409 = 0.000273891$$

2.b Filtering

$$P(X6 = S_1|O = ACCGTA;\Theta) = \frac{\alpha_6(1)\beta_6(1)}{\alpha_6(1)\beta_6(1) + \alpha_6(2)\beta_6(2)}$$

We know that, for the final time stamp $t = 6$, $\beta_6(1) = \beta_6(2) = 1$

$$P(X6 = S_1|O = ACCGTA;\Theta) = \frac{0.000184482}{0.000184482 + 0.000089409} = 0.6735599$$

$$P(X6 = S_2|O = ACCGTA;\Theta) = \frac{0.000089409}{0.000184482 + 0.000089409} = 0.32644$$

2.c Smoothing

t	O_t	$\beta_t(s_1)$	$\beta_t(s_2)$
6	A	$\beta_6(1) = 1$	$\beta_6(2) = 1$
5	T	$\beta_5(1) =$ $a_{11}P(A\|X_6 = s_1)\beta_6(1) + a_{12}P(A\|X_6 = s_2)\beta_6(2)$ $= 0.7 \times 0.4 \times 1 + 0.3 \times 0.2 \times 1$ $= 0.34$	$\beta_5(2) =$ $a_{21}P(A\|X_6 = s_1)\beta_6(1) + a_{22}P(A\|X_6 = s_2)\beta_6(2)$ $= 0.4 \times 0.4 \times 1 + 0.6 \times 0.2 \times 1$ $= 0.28$
4	G	$\beta_4(1) =$ $a_{11}P(T\|X_5 = s_1)\beta_5(1) + a_{12}P(T\|X_5 = s_2)\beta_(2)$ $= 0.7 \times 0.1 \times 0.34 + 0.3 \times 0.3 \times 0.28$ $= 0.049$	$\beta_4(2) =$ $a_{21}P(T\|X_6 = s_1)\beta_6(1) + a_{22}P(T\|X_6 = s_2)\beta_6(2)$ $= 0.4 \times 0.1 \times 0.34 + 0.6 \times 0.3 \times 0.28$ $= 0.064$

$$\begin{aligned} P(X4 = s_1|O = ACCGTA;\Theta) &= \frac{\alpha_4(1)\beta_4(1)}{\alpha_4(1)\beta_4(1) + \alpha_4(2)\beta_4(2)} \\ &= \frac{0.0039408 \times 0.049}{0.0039408 \times 0.049 + 0.0012624 \times 0.064} \\ &= 0.1928995 \end{aligned}$$

$$\begin{aligned} P(X4 = s_2|O = ACCGTA;\Theta) &= \frac{\alpha_4(2)\beta_4(2)}{\alpha_4(1)\beta_4(1) + \alpha_4(2)\beta_4(2)} \\ &= \frac{0.0012624 \times 0.064}{0.0039408 \times 0.049 + 0.0012624 \times 0.064} \\ &= 0.8071 \end{aligned}$$

2.d Most likely explanation

We know that the recursive definition of Viterbi algorithm for maximum likehood path in HMM is given by:

$$\delta_t(j) = \max_i \delta_{t-1}(i) a_{ij} P(o_t|X_t = s_j) \tag{2}$$

$$\begin{aligned}
\delta_1(1) &= 0.6 \times 0.4 = 0.24 \\
\delta_1(2) &= 0.4 \times 0.2 = 0.08 \\
\delta_2(1) &= \max\{\delta_1(1)a_{11}, \delta_1(2)a_{21}\} \times P(C|X_t = s_1) \\
&= \max\{0.24 \times 0.7, 0.08 \times 0.4\} \times 0.2 = 0.0336 \\
\delta_2(2) &= \max\{\delta_1(1)a_{12}, \delta_1(2)a_{22}\} \times P(C|X_t = s_2) \\
&= \max\{0.24 \times 0.3, 0.08 \times 0.6\} \times 0.4 = 0.0288 \\
\delta_3(1) &= \max\{\delta_2(1)a_{11}, \delta_2(2)a_{21}\} \times P(C|X_t = s_1) \\
&= \max\{0.0336 \times 0.7, 0.0288 \times 0.4\} \times 0.2 = 0.004704 \\
\delta_3(2) &= \max\{\delta_2(1)a_{12}, \delta_2(2)a_{22}\} \times P(C|X_t = s_2) \\
&= \max\{0.0336 \times 0.3, 0.0288 \times 0.6\} \times 0.4 = 0.003456 \\
\delta_4(1) &= \max\{\delta_3(1)a_{11}, \delta_3(2)a_{21}\} \times P(G|X_t = s_1) \\
&= \max\{0.004704 \times 0.7, 0.003456 \times 0.4\} \times 0.2 = 0.00065856 \\
\delta_4(2) &= \max\{\delta_3(1)a_{12}, \delta_3(2)a_{22}\} \times P(G|X_t = s_2) \\
&= \max\{0.004704 \times 0.3, 0.003456 \times 0.6\} \times 0.4 = 0.00082944 \\
\delta_5(1) &= \max\{\delta_4(1)a_{11}, \delta_4(2)a_{21}\} \times P(T|X_t = s_1) \\
&= \max\{0.00065856 \times 0.7, 0.00082944 \times 0.4\} \times 0.1 = 0.0000460992 \\
\delta_5(2) &= \max\{\delta_4(1)a_{12}, \delta_4(2)a_{22}\} \times P(T|X_t = s_2) \\
&= \max\{0.00065856 \times 0.3, 0.00082944 \times 0.6\} \times 0.3 = 0.0001492992 \\
\delta_6(1) &= \max\{\delta_5(1)a_{11}, \delta_5(2)a_{21}\} \times P(A|X_t = s_1) \\
&= \max\{0.0000460992 \times 0.7, 0.0001492992 \times 0.4\} \times 0.4 = 0.000023887872 \\
\delta_6(2) &= \max\{\delta_5(1)a_{12}, \delta_5(2)a_{22}\} \times P(A|X_t = s_2) \\
&= \max\{0.0000460992 \times 0.3, 0.0001492992 \times 0.6\} \times 0.2 = 0.000017915904 \\
&\text{//TODO: complete the calculations like above}
\end{aligned}$$

So, the most likely explanation for the states that produced the observations **ACCGTA** $= s_1 s_1 s_1 s_2 s_2 s_1$

2.e Prediction

$$\begin{aligned}
P(O_7|O) &= \sum_{j \in 1,2} P(O_7, X_7 = s_j) \\
&= \sum_{j \in 1,2} P(O_7|X_7 = s_j) P(X_7 = s_j, O) \\
&= \sum_{j \in 1,2} P(O_7|X_7 = s_j) \times \sum_{k \in 1,2} P(X_7 = s_j|X_6 = s_k) P(X_6 = s_k, O) \\
&= P(O_7|X_7 = s_1) \times \sum_{k \in 1,2} P(X_7 = s_1|X_6 = s_k) P(X_6 = s_k, O) \\
&\quad + P(O_7|X_7 = s_2) \times \sum_{k \in 1,2} P(X_7 = s_2|X_6 = s_k) P(X_6 = s_k, O) \\
&= P(O_7|X_7 = s_1)[P(X_7 = s_1|X_6 = s_1)P(X_6 = s_1, O) + P(X_7 = s_1|X_6 = s_2)P(X_6 = s_2, O)] \\
&\quad + P(O_7|X_7 = s_2)[P(X_7 = s_2|X_6 = s_1)P(X_6 = s_1, O) + P(X_7 = s_2|X_6 = s_2)P(X_6 = s_2, O)] \\
&= P(O_7|X_7 = s_1)[0.7 \times 0.6735599 + 0.4 \times 0.32644] + P(O_7|X_7 = s_2)[0.3 \times 0.6735599 + 0.6 \times 0.32644)] \\
P(O_7|O) &= 0.60206793 \times P(O_7|X_7 = s_1) + 0.39793197 \times P(O_7|X_7 = s_2)
\end{aligned} \tag{3}$$

$$
\begin{aligned}
P(O_7 = A|O) &= 0.60206793 \times 0.4 + 0.39793197 \times 0.2 = 0.320413566 \\
P(O_7 = C|O) &= 0.60206793 \times 0.2 + 0.39793197 \times 0.4 = 0.279586374 \\
P(O_7 = G|O) &= 0.60206793 \times 0.3 + 0.39793197 \times 0.1 = 0.220413576 \\
P(O_7 = T|O) &= 0.60206793 \times 0.1 + 0.39793197 \times 0.3 = 0.179586384
\end{aligned}
$$

So, our prediction for O_7, the next observation, is **'A'**.