

# CSCI 567–HOMEWORK 5

Chengliang Dong (Meng Liu, Kan Wang, Ye Feng and Peehoo Dewan)

Monday 1<sup>st</sup> December, 2014

## 1 Principal Component Analysis

### 1.1 Minimum reconstruction error

#### 1.1.1 When $u$ is fixed

$$\begin{aligned} L_{re} &= \sum_{i=1}^N (x_i - uz_i)^T (x_i - uz_i) \\ \frac{\partial L_{re}}{\partial z_i} &= \nabla_{z_i} \text{tr} \left( \sum_{i=1}^N (x_i - uz_i)^T (x_i - uz_i) \right) \\ &= -2u^T x_i + 2u^T uz_i \\ &= 0 \\ z_i^* &= u^T x_i \end{aligned}$$

#### 1.1.2 Derive optimal $u$

$$\begin{aligned} L_u &= \sum_{i=1}^N (x_i - uz_i)^T (x_i - uz_i) + \lambda(u^T u - I) \\ \frac{\partial L_u}{\partial u} &= \nabla_u \text{tr} \left( \sum_{i=1}^N (x_i - uz_i)^T (x_i - uz_i) \right) + 2\lambda u \\ &= -2 \sum_{i=1}^N x_i x_i^T u + 2\lambda u \\ &= 0 \\ \sum_{i=1}^N x_i x_i^T u^* &= \lambda u^* \end{aligned}$$

According to the definition of eigenvector,  $u^*$  is the eigenvectors of covariance matrix.

## 1.2 Gaussian distribution

### 1.2.1 Optimal $p$

According to the property of Gaussian distribution,  $z \sim N(0, p^T \Sigma p)$ . Therefore, we have,

$$H(z) = \frac{1}{2}(1 + \ln 2\pi) + \frac{1}{2} \ln(p^T \Sigma p)$$

Therefore, maximizing entropy of  $z$  is the same as maximizing  $\ln(p^T \Sigma p)$ . As logarithm is an increasing function, then maximizing  $\ln(p^T \Sigma p)$  is the same as maximizing  $p^T \Sigma p$ . So the problem is reduced to

maximizing  $p^T \Sigma p$  with the constraint that  $p^T p = I$ .

$$\begin{aligned}
 L_p &= p^T \Sigma p + \lambda(p^T p - I) \\
 \frac{\partial L_p}{\partial p} &= \nabla_p \text{tr}(p^T \Sigma p) + 2\lambda p \\
 &= -2\Sigma p + 2\lambda p \\
 &= 0 \\
 \Sigma p^* &= \lambda p^*
 \end{aligned}$$

### 1.2.2 Maximizing variance

From last question, we see  $p^*$  is the eigenvector of covariance matrix  $\Sigma$ . Therefore, obviously, it is also the solution of PCA, which can be derived from maximizing variance of  $z$ .

## 2 Hidden Markov Models

### 2.1 $P(e|\theta)$

$$\alpha_1(1) = \pi_1 b_1(C) = 0.5 \times 0.1 = 0.05$$

$$\alpha_1(2) = \pi_1 b_1(C) = 0.5 \times 0.4 = 0.20$$

$$\alpha_2(1) = (\alpha_1(1)a_{11} + \alpha_1(2)a_{21})b_1(G) = (0.05 \times 0.7 + 0.20 \times 0.3) \times 0.4 = 0.038$$

$$\alpha_2(2) = (\alpha_1(2)a_{12} + \alpha_1(2)a_{22})b_2(G) = (0.20 \times 0.7 + 0.05 \times 0.3) \times 0.1 = 0.0155$$

$$\alpha_3(1) = (\alpha_2(1)a_{11} + \alpha_2(2)a_{21})b_1(T) = (0.038 \times 0.7 + 0.0155 \times 0.3) \times 0.1 = 0.003125$$

$$\alpha_3(2) = (\alpha_2(2)a_{12} + \alpha_2(2)a_{22})b_2(T) = (0.0155 \times 0.7 + 0.038 \times 0.3) \times 0.4 = 0.0089$$

$$\alpha_4(1) = (\alpha_3(1)a_{11} + \alpha_3(2)a_{21})b_1(C) = (0.003125 \times 0.7 + 0.0089 \times 0.3) \times 0.1 = 0.00048575$$

$$\alpha_4(2) = (\alpha_3(2)a_{12} + \alpha_3(2)a_{22})b_2(C) = (0.0089 \times 0.7 + 0.0089 \times 0.3) \times 0.4 = 0.002867$$

$$\alpha_5(1) = (\alpha_4(1)a_{11} + \alpha_4(2)a_{21})b_1(A) = (0.0004857 \times 0.7 + 0.002867 \times 0.3) \times 0.4 = 0.00048005$$

$$\alpha_5(2) = (\alpha_4(2)a_{12} + \alpha_4(2)a_{22})b_2(A) = (0.002867 \times 0.7 + 0.0004857 \times 0.3) \times 0.1 = 0.0002152625$$

$$\alpha_6(1) = (\alpha_5(1)a_{11} + \alpha_5(2)a_{21})b_1(A) = (0.00048005 \times 0.7 + 0.0002152625 \times 0.3) \times 0.4 = 0.0001602455$$

$$\alpha_6(2) = (\alpha_5(2)a_{12} + \alpha_5(2)a_{22})b_2(A) = (0.0002152625 \times 0.7 + 0.00048005 \times 0.3) \times 0.1 = 0.00002946987$$

$$P(e|\theta) = \alpha_6(1) + \alpha_6(2) = 0.0001897154$$

## 2.2 $P(X_t = S_i|e, \theta)$

$$\beta_6(1) = 1$$

$$\beta_6(2) = 1$$

$$\beta_5(1) = (a_{11}b_1(G)\beta_6(1) + a_{12}b_2(G)\beta_6(2)) = (0.7 \times 0.4 \times 1 + 0.3 \times 0.1 \times 1) = 0.31$$

$$\beta_5(2) = (a_{21}b_1(G)\beta_6(2) + a_{22}b_2(G)\beta_6(1)) = (0.3 \times 0.4 \times 1 + 0.7 \times 0.1 \times 1) = 0.19$$

$$\beta_4(1) = (a_{11}b_1(A)\beta_5(1) + a_{12}b_2(A)\beta_5(2)) = (0.7 \times 0.4 \times 0.31 + 0.3 \times 0.1 \times 0.19) = 0.0925$$

$$\beta_4(2) = (a_{21}b_1(A)\beta_5(2) + a_{22}b_2(A)\beta_5(1)) = (0.3 \times 0.4 \times 0.19 + 0.7 \times 0.1 \times 0.31) = 0.0505$$

$$\beta_3(1) = (a_{11}b_1(C)\beta_4(1) + a_{12}b_2(C)\beta_4(2)) = (0.7 \times 0.1 \times 0.0925 + 0.3 \times 0.4 \times 0.0505) = 0.012535$$

$$\beta_3(2) = (a_{21}b_1(C)\beta_4(2) + a_{22}b_2(C)\beta_4(1)) = (0.3 \times 0.1 \times 0.0505 + 0.7 \times 0.4 \times 0.0925) = 0.016915$$

$$\beta_2(1) = (a_{11}b_1(T)\beta_3(1) + a_{12}b_2(T)\beta_3(2)) = (0.7 \times 0.1 \times 0.012535 + 0.3 \times 0.4 \times 0.016915) = 0.00290725$$

$$\beta_2(2) = (a_{21}b_1(T)\beta_3(2) + a_{22}b_2(T)\beta_3(1)) = (0.3 \times 0.1 \times 0.016915 + 0.7 \times 0.4 \times 0.012535) = 0.00511225$$

$$\begin{aligned}\beta_1(1) &= (a_{11}b_1(G)\beta_2(1) + a_{12}b_2(G)\beta_2(2)) = (0.7 \times 0.4 \times 0.00290725 + 0.3 \times 0.1 \times 0.00511225) \\ &= 0.0009673975\end{aligned}$$

$$\begin{aligned}\beta_1(2) &= (a_{21}b_1(G)\beta_2(2) + a_{22}b_2(G)\beta_2(1)) = (0.3 \times 0.4 \times 0.00511225 + 0.7 \times 0.1 \times 0.00290725) \\ &= 0.0007067275\end{aligned}$$

Therefore, we have,

$$P(X_1 = S_1|e, \theta) = \frac{\alpha_1(1)\beta_1(1)}{\alpha_1(1)\beta_1(1) + \alpha_1(2)\beta_1(2)} = \frac{0.05 \times 0.0009673975}{0.05 \times 0.0009673975 + 0.20 \times 0.0007067275} = 0.2549602$$

$$P(X_1 = S_2|e, \theta) = 1 - P(X_1 = S_1|e, \theta) = 0.7450398$$

$$P(X_2 = S_1|e, \theta) = \frac{\alpha_2(1)\beta_2(1)}{\alpha_2(1)\beta_2(1) + \alpha_2(2)\beta_2(2)} = \frac{0.038 \times 0.00290725}{0.038 \times 0.00290725 + 0.0155 \times 0.00511225} = 0.5823223$$

$$P(X_2 = S_2|e, \theta) = 1 - P(X_2 = S_1|e, \theta) = 0.4176777$$

$$P(X_3 = S_1|e, \theta) = \frac{\alpha_3(1)\beta_3(1)}{\alpha_3(1)\beta_3(1) + \alpha_3(2)\beta_3(2)} = \frac{0.003125 \times 0.012535}{0.003125 \times 0.012535 + 0.0089 \times 0.016915} = 0.2064771$$

$$P(X_3 = S_2|e, \theta) = 1 - P(X_3 = S_1|e, \theta) = 0.7935229$$

$$P(X_4 = S_1|e, \theta) = \frac{\alpha_4(1)\beta_4(1)}{\alpha_4(1)\beta_4(1) + \alpha_4(2)\beta_4(2)} = \frac{0.00048575 \times 0.0925}{0.00048575 \times 0.0925 + 0.002867 \times 0.0505} = 0.2368383$$

$$P(X_4 = S_2|e, \theta) = 1 - P(X_4 = S_1|e, \theta) = 0.7631617$$

$$P(X_5 = S_1|e, \theta) = \frac{\alpha_5(1)\beta_5(1)}{\alpha_5(1)\beta_5(1) + \alpha_5(2)\beta_5(2)} = \frac{0.00048005 \times 0.31}{0.00048005 \times 0.31 + 0.0002152625 \times 0.19} = 0.7844145$$

$$P(X_5 = S_2|e, \theta) = 1 - P(X_5 = S_1|e, \theta) = 0.2155855$$

$$P(X_6 = S_1|e, \theta) = \frac{\alpha_6(1)\beta_6(1)}{\alpha_6(1)\beta_6(1) + \alpha_6(2)\beta_6(2)} = \frac{0.0001602455 \times 1}{0.0001602455 \times 1 + 0.00002946987 \times 1} = 0.8446627$$

$$P(X_6 = S_2|e, \theta) = 1 - P(X_6 = S_1|e, \theta) = 0.1553373$$

## 2.3 Viterbi algorithm

*route* :  $S_1$

$$\delta_1(1) = \pi_1 b_1(C) = 0.5 \times 0.1 = 0.05$$

*route* :  $S_2$

$$\delta_1(2) = \pi_2 b_2(C) = 0.5 \times 0.4 = 0.20$$

*route* :  $S_2 S_1$

$$\delta_2(1) = \max(\delta_1(1)a_{11}, \delta_1(2)a_{21})b_1(G) = 0.055 \times 0.4 = 0.024$$

*route* :  $S_2 S_2$

$$\delta_2(2) = \max(\delta_1(1)a_{12}, \delta_1(2)a_{22})b_2(G) = 0.14 \times 0.1 = 0.014$$

*route* :  $S_2 S_1 S_1$

$$\delta_3(1) = \max(\delta_2(1)a_{11}, \delta_2(2)a_{21})b_1(T) = 0.0168 \times 0.1 = 0.00168$$

*route* :  $S_2 S_2 S_2$

$$\delta_3(2) = \max(\delta_2(1)a_{12}, \delta_2(2)a_{22})b_2(T) = 0.0098 \times 0.4 = 0.00392$$

*route* :  $S_2 S_1 S_1 S_1$

$$\begin{aligned} \delta_4(1) &= \max(\delta_3(1)a_{11}, \delta_3(2)a_{21})b_1(C) = 0.001176 \times 0.1 \\ &= 0.0001176 \end{aligned}$$

*route* :  $S_2 S_2 S_2 S_2$

$$\begin{aligned} \delta_4(2) &= \max(\delta_3(1)a_{12}, \delta_3(2)a_{22})b_2(C) = 0.002744 \times 0.4 \\ &= 0.0010976 \end{aligned}$$

*route* :  $S_2 S_2 S_2 S_2 S_1$

$$\begin{aligned} \delta_5(1) &= \max(\delta_4(1)a_{11}, \delta_4(2)a_{21})b_1(A) = 0.00032928 \times 0.4 \\ &= 0.000131712 \end{aligned}$$

*route* :  $S_2 S_2 S_2 S_2 S_2$

$$\begin{aligned} \delta_5(2) &= \max(\delta_4(1)a_{12}, \delta_4(2)a_{22})b_2(A) = 0.00076832 \times 0.1 \\ &= 0.000076832 \end{aligned}$$

*route* :  $S_2 S_2 S_2 S_2 S_1 S_1$

$$\begin{aligned} \delta_6(1) &= \max(\delta_5(1)a_{11}, \delta_5(2)a_{21})b_1(G) = 0.0000921984 \times 0.4 \\ &= 0.00003687936 \end{aligned}$$

*route* :  $S_2 S_2 S_2 S_2 S_2 S_2$

$$\begin{aligned} \delta_6(2) &= \max(\delta_5(1)a_{12}, \delta_5(2)a_{22})b_2(G) = 0.0000537824 \times 0.1 \\ &= 0.00000537824 \end{aligned}$$

Therefore, the most probable route is  $S_2 S_2 S_2 S_2 S_1 S_1$ , while the sequence of most likely states estimated independently is  $S_2 S_1 S_2 S_2 S_1 S_1$ . They are not the same.

## 2.4 Prediction

$$\begin{aligned}
P(e_7|e) &= \sum_{i=1}^2 P(e_7, S_i|e) \\
&= \sum_{i=1}^2 P(e_7|S_i)P(S_i|e) \\
&= \sum_{i=1}^2 P(e_7|S_i) \sum_{j=1}^2 P(S_i, S_j|e) \\
&= \sum_{i=1}^2 P(e_7|S_i) \sum_{j=1}^2 P(S_i|S_j)P(S_j|e) \\
&= \frac{1}{P(e)} \sum_{i=1}^2 P(e_7|S_i) \sum_{j=1}^2 P(S_i|S_j)P(S_j, e) \\
&= C \sum_{i=1}^2 P(e_7|S_i) \sum_{j=1}^2 P(S_i|S_j)P(S_j, e) \\
&= C \sum_{i=1}^2 P(e_7|S_i) \sum_{j=1}^2 P(S_i|S_j)\alpha_6(j)
\end{aligned}$$

Where C is a constant. Therefore,

If  $e_7 = A$ , then we have,

$$\begin{aligned}
P(A|e) &= C(P(A|S_1) \sum_{j=1}^2 P(S_1|S_j)\alpha_6(j) + P(A|S_2) \sum_{j=1}^2 P(S_2|S_j)\alpha_6(j)) \\
&= C(P(A|S_1)(P(S_1|S_1)\alpha_6(1) + P(S_1|S_2)\alpha_6(2)) + P(A|S_2)(P(S_2|S_1)\alpha_6(1) + P(S_2|S_2)\alpha_6(2))) \\
&= C \times 0.00005527538
\end{aligned}$$

$$\begin{aligned}
P(T|e) &= C(P(T|S_1) \sum_{j=1}^2 P(S_1|S_j)\alpha_6(j) + P(T|S_2) \sum_{j=1}^2 P(S_2|S_j)\alpha_6(j)) \\
&= C(P(T|S_1)(P(S_1|S_1)\alpha_6(1) + P(S_1|S_2)\alpha_6(2)) + P(T|S_2)(P(S_2|S_1)\alpha_6(1) + P(S_2|S_2)\alpha_6(2))) \\
&= C \times 0.0000395823
\end{aligned}$$

$$\begin{aligned}
P(C|e) &= C(P(C|S_1) \sum_{j=1}^2 P(S_1|S_j)\alpha_6(j) + P(C|S_2) \sum_{j=1}^2 P(S_2|S_j)\alpha_6(j)) \\
&= C(P(C|S_1)(P(S_1|S_1)\alpha_6(1) + P(S_1|S_2)\alpha_6(2)) + P(C|S_2)(P(S_2|S_1)\alpha_6(1) + P(S_2|S_2)\alpha_6(2))) \\
&= C \times 0.0000395823
\end{aligned}$$

$$\begin{aligned}
P(G|e) &= C(P(G|S_1) \sum_{j=1}^2 P(S_1|S_j)\alpha_6(j) + P(G|S_2) \sum_{j=1}^2 P(S_2|S_j)\alpha_6(j)) \\
&= C(P(G|S_1)(P(S_1|S_1)\alpha_6(1) + P(S_1|S_2)\alpha_6(2)) + P(G|S_2)(P(S_2|S_1)\alpha_6(1) + P(S_2|S_2)\alpha_6(2))) \\
&= C \times 0.00005527538
\end{aligned}$$

Therefore, it could be either A or G.

### 3 Programming

#### 3.1 Implementing PCA

Done.

#### 3.2 Obtain eigenfaces

Figure 1.



(a) first eigenface



(b) second eigenface



(c) third eigenface



(d) fourth eigenface



(e) fifth eigenface

Figure 1: First five Eigenfaces from PCA

### 3.3 Classification

#### 3.3.1 Linear SVM cross-validation results

Accuracy (in percentage) for linear SVM at different cost for different dimensions.

C	d=20	d=50	d=100	d=200
$4^{-6}$	25.9375	29.0625	29.0625	28.7500
$4^{-5}$	34.5312	39.2188	39.8438	39.6875
$4^{-4}$	62.0312	71.2500	73.2812	73.2812
$4^{-3}$	79.0625	87.8125	91.2500	92.0312
$4^{-2}$	86.4062	92.9688	94.5312	95.0000
$4^{-1}$	90.6250	94.2188	95.1562	96.2500
$4^0$	93.7500	95.4688	95.3125	95.4688
$4^1$	92.9688	93.5938	94.6875	95.3125
$4^2$	89.6875	92.5000	93.5938	94.6875

Therefore, we have following tuned parameter for linear SVM. For d=20, cost= $4^1$  is the optimal parameter. For d=50, cost=1 is the optimal parameter. For d=100, cost=1 is the optimal parameter. For d=200, cost= $4^{-1}$  is the optimal parameter.

#### 3.3.2 radial SVM cross-validation results

Accuracy (in percentage) for radial SVM at different cost for d=20.

C	gamma= $4^{-7}$	gamma= $4^{-6}$	gamma= $4^{-5}$	gamma= $4^{-4}$	gamma= $4^{-3}$	gamma= $4^{-2}$	gamma= $4^{-1}$
$4^{-6}$	26.0938	27.0312	30.9375	35.3125	47.3438	40.7812	24.0625
$4^{-5}$	26.0938	27.0312	30.9375	35.3125	47.3438	40.7812	24.0625
$4^{-4}$	26.0938	27.0312	30.9375	35.3125	47.3438	40.7812	24.0625
$4^{-3}$	26.0938	27.0312	30.9375	35.3125	47.3438	40.7812	24.0625
$4^{-2}$	26.0938	27.0312	30.9375	35.3125	47.3438	40.7812	24.0625
$4^{-1}$	26.0938	27.0312	30.9375	41.5625	51.4062	41.7188	24.0625
$4^0$	26.0938	29.2188	48.2812	63.9062	70.4688	62.8125	41.0938
$4^1$	29.0625	49.6875	69.2188	81.8750	82.9688	71.2500	43.1250
$4^2$	49.8438	70.1562	82.1875	89.0625	87.9688	71.5625	43.5938

Accuracy (in percentage) for radial SVM at different cost for d=50.

C	$\gamma=4^{-7}$	$\gamma=4^{-6}$	$\gamma=4^{-5}$	$\gamma=4^{-4}$	$\gamma=4^{-3}$	$\gamma=4^{-2}$	$\gamma=4^{-1}$
$4^{-6}$	29.2188	30.3125	35.3125	42.5000	55.7812	43.5938	24.3750
$4^{-5}$	29.2188	30.3125	35.3125	42.5000	55.7812	43.5938	24.3750
$4^{-4}$	29.2188	30.3125	35.3125	42.5000	55.7812	43.5938	24.3750
$4^{-3}$	29.2188	30.3125	35.3125	42.5000	55.7812	43.5938	24.3750
$4^{-2}$	29.2188	30.3125	35.3125	42.5000	55.7812	43.5938	24.3750
$4^{-1}$	29.2188	30.3125	35.4688	49.8438	61.8750	44.5312	24.3750
$4^0$	29.2188	32.9688	55.1562	77.5000	84.6875	76.8750	42.8125
$4^1$	32.3438	55.1562	79.0625	90.4688	90.4688	79.2188	45.3125
$4^2$	55.3125	79.6875	90.3125	93.7500	93.2812	79.5312	45.4688

Accuracy (in percentage) for radial SVM at different cost for d=100.

C	$\gamma=4^{-7}$	$\gamma=4^{-6}$	$\gamma=4^{-5}$	$\gamma=4^{-4}$	$\gamma=4^{-3}$	$\gamma=4^{-2}$	$\gamma=4^{-1}$
$4^{-6}$	29.2188	30.0000	35.6250	43.1250	56.0938	40.6250	24.0625
$4^{-5}$	29.2188	30.0000	35.6250	43.1250	56.0938	40.6250	24.0625
$4^{-4}$	29.2188	30.0000	35.6250	43.1250	56.0938	40.6250	24.0625
$4^{-3}$	29.2188	30.0000	35.6250	43.1250	56.0938	40.6250	24.0625
$4^{-2}$	29.2188	30.0000	35.6250	43.1250	56.0938	40.6250	24.0625
$4^{-1}$	29.2188	30.0000	35.7812	50.1562	63.1250	40.6250	24.0625
$4^0$	29.2188	32.8125	56.2500	79.3750	86.5625	77.1875	41.2500
$4^1$	32.3438	56.5625	82.6562	92.9688	92.1875	79.2188	43.7500
$4^2$	55.7812	83.4375	93.1250	94.5312	93.4375	79.0625	43.9062

Accuracy (in percentage) for radial SVM at different cost for d=200.

C	$\gamma=4^{-7}$	$\gamma=4^{-6}$	$\gamma=4^{-5}$	$\gamma=4^{-4}$	$\gamma=4^{-3}$	$\gamma=4^{-2}$	$\gamma=4^{-1}$
$4^{-6}$	28.7500	30.0000	35.4688	43.4375	55.7812	39.6875	24.6875
$4^{-5}$	28.7500	30.0000	35.4688	43.4375	55.7812	39.6875	24.6875
$4^{-4}$	28.7500	30.0000	35.4688	43.4375	55.7812	39.6875	24.6875
$4^{-3}$	28.7500	30.0000	35.4688	43.4375	55.7812	39.6875	24.6875
$4^{-2}$	28.7500	30.0000	35.4688	43.4375	55.7812	39.6875	24.6875
$4^{-1}$	28.7500	30.0000	35.6250	50.4688	62.8125	39.6875	24.6875
$4^0$	28.7500	32.8125	56.2500	79.2188	86.8750	75.6250	39.3750
$4^1$	31.8750	56.2500	82.9688	92.9688	92.5000	77.8125	42.3438
$4^2$	55.7812	83.7500	93.2812	95.1562	93.4375	78.4375	42.3438

Therefore, we have following tuned parameter for radial SVM. For d=20, 50, 100, 200, cost= $4^2$   $\gamma=4^{-4}$  is always the optimal set of parameters.



## 3.4 HMM

### 3.4.1 Implementing HMM

Done.

### 3.4.2 Parameter estimation

From implemented HMM model with initial  $pi_1 = 0.1$  and  $pi_2 = 0.9$ , after 500 iterations, we obtained such estimate for A and E.

$$\mathbf{A} = \begin{pmatrix} 0.9171 & 0.0829 \\ 0.0729 & 0.9271 \end{pmatrix}$$
$$\mathbf{E} = \begin{pmatrix} 0.0979 & 0.4303 & 0.3924 & 0.0794 \\ 0.3868 & 0.1023 & 0.1146 & 0.3963 \end{pmatrix}$$

On the other hand, from hmmtrain package, the estimate A is,

$$\mathbf{A} = \begin{pmatrix} 0.9249 & 0.0751 \\ 0.0844 & 0.9156 \end{pmatrix} \quad (-1)$$

They are similar results, with differences probably owing to initialization and the number of iterations.