REVIEW EXERCISES
Math347, Fall 2010

1. We have two univariate linear regression models:

$$y_{j1} = \beta_{01} + \beta_{11} z_{j1} + \epsilon_{j1},$$
$$y_{j2} = \beta_{02} + \beta_{12} z_{j2} + \epsilon_{j2}, \qquad j = 1, 2, 3, 4, 5.$$

Here $(\epsilon_{j1}, \epsilon_{j2})$ are iid following two dimensional normal distribution with mean 0 and variance $\Sigma_{2\times 2}$. The five observations are

z:	1	-1	0	2	-2
y_1:	0	1	2	-1	0
y_2:	2	1	-1	0	-2

a. Give the least squares estimator of $\begin{pmatrix} \beta_{01} & \beta_{11} \\ \beta_{02} & \beta_{12} \end{pmatrix}$.

b. Find the residuals $(\hat{\epsilon}_{j1}, \hat{\epsilon}_{j2}), 1 \le j \le 5$.

c. Find the MLE $\hat{\Sigma}$ of Σ and the unbiased estimator $\hat{\mathbf{S}}$ of Σ.

2. Over the past year of 52 weeks, the weekly rate of return of stock market (x_1) and the weekly rate of return of real estate market (x_2) are monitored. The records show that the sample correlation matrix of $R = cor\binom{x_1}{x_2}$

a. $R = \begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix}$, find the two principal components.

b. $R = \begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$, find the two principal components.

c. How many principal components would you like to retain in part a and part b? Give some explanation based on the size of correlations.

3. We are to evaluate the economic performance of Hong Kong in the past few years. In our dataset we have observations of three variables: x_1: inflation, x_2: unemployment, x_3: GDP. Suppose we shall use orthogonal factor model to analyze the data. The eigenvalues and eigenvectors of the sample covariance matrix of three variables are

$$\hat{\lambda}_1 = 2.25 \qquad \hat{\lambda}_2 = 1.96 \qquad \hat{\lambda}_3 = 0.16$$

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{pmatrix} \qquad \hat{\mathbf{e}}_2 = \begin{pmatrix} 1/2 \\ 1/2 \\ -1/\sqrt{2} \end{pmatrix} \qquad \hat{\mathbf{e}}_3 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}$$

a. Reconstruct the sample covariance matrix based on the above eigenvalues and eigenvectors.

b. Assume there is only one factor in the orthogonal factor model (i.e, $m = 1$). Use principal component approach to find 1). factor loading matrix, 2). communalities, 3). specific variances, 4) residual matrix, 5). proportion of variances explained by the first factor.

c. Repeat part b assuming $m = 2$. (For question 5, answer the proportion of variances explained by the two factors.) Do you prefer to using the model with $m = 1$ in part a or using the model with $m = 2$ in part b? Explain why.

d. For model with $m = 2$ in part b, plot the three variables (inflation, unemployment and GDP) with the factor loadings as the coordinates and the two factors as axis. You must indicate which point is which variable in your plot.

e. Following part d, rotate the factors so that, after rotation, GDP has factor loading 0 on the second factor and positive factor loading on the first factor. Compute all factor loadings for this rotated factor model. And provide interpretation for the rotated factors.

4. X_1 are X_2 are two one-dimensional normal random variables: $X_1 \sim N(1, 1)$ and $X_2 \sim N(0, 4)$. Suppose the correlation of X_1 and X_2 is 0.5.

a). Write down the variance matrix of $\mathbf{X} \equiv \binom{X_1}{X_2}$, the density function of X_1 and the density function of X_2. Do you know the joint density function of $\mathbf{X}$?

b). Find the principal components of $\mathbf{X}$ based on the variance matrix.

c). Find the two two-dimensional unit vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ such that $\|\mathbf{b}_1\| = \|\mathbf{b}_2\| = 1$ and

$$Var(\mathbf{b}_1'\mathbf{X}) = \max_{\mathbf{b}:\|\mathbf{b}\|=1} \{Var(\mathbf{b}'\mathbf{X})\}$$
$$Var(\mathbf{b}_2'\mathbf{X}) = \min_{\mathbf{b}:\|\mathbf{b}\|=1} \{Var(\mathbf{b}'\mathbf{X})\}.$$

Here $\| \cdot \|$ is the Euclidean norm.

5. The eigenvalues and eigenvectors of a 3×3 variance matrix of $X = (x_1, x_2, x_3)'$ are

$$\lambda_1 = 12 \qquad \mathbf{e}_1' = (1,\ -1,\ -1)/\sqrt{3}$$
$$\lambda_2 = 6 \qquad \mathbf{e}_2' = (2,\ 1,\ 1)/\sqrt{6}$$
$$\lambda_3 = .25 \qquad \mathbf{e}_3' = (0,\ -1,\ 1)/\sqrt{2}$$

a). Assume $m = 2$, calculate the factor loading matrix L by principal component approach.

b). Plot the factor loadings with the two factors F_1 and F_2 being the two horizontal and vertical axes.

c). Rotate the factors (F_1, F_2) anticlockwise to (F_1^*, F_2^*) so that the third variable x_3 has zero loading onto F_1^*. Compute the rotated factor loadings and plot the rotated factors in your figure in part b).

6. Let z_1, z_2, z_3 and z_4 be four indenpendent one dimensional standard normal random variables. Let $\mathbf{X} = \binom{\mathbf{X}_1}{\mathbf{X}_2}$ with $\mathbf{X}_1 = \binom{x_{11}}{x_{12}}$ and $\mathbf{X}_2 = \binom{x_{21}}{x_{22}}$ be such that

$$x_{11} = z_1, \qquad x_{12} = z_2,$$
$$x_{21} = z_1 + z_3, \qquad x_{22} = z_1 + z_2 - z_3 + z_4.$$

a). Compute the sample correlation matrix $\mathbf{R}$ of $\mathbf{X}$.

b). Based on $\mathbf{R}$, compute the canonical correlations ρ_1^*, ρ_2^* and the canonical variates (U_1, V_1) and (U_2, V_2).

c). Following part b), compute the "percentage of variance of $\mathbf{X}_1$ explained by U_1" and the "percentage of variance of $\mathbf{X}_2$ explained by V_1".

7. Three randomly chosen graduates from HKU and three randomly chosen graduates from HKUST are tested on math and literature problems, their scores are listed in the following

Population 1		Population 2	
HKU		HKUST	
math	literature	math	literature
10	15	10	20
20	10	30	30
15	10	20	25

Assume the scores of graduates from the two universities follow two bivariate normal distribution with equal variances. In addition, assume the students are randomly chosen from their universities.

a) Suppose we can assume $p_1 = p_2$. Establish a classification rule based on the above six observations, assuming equal costs of misclassfication $c(1|2) = c(2|1)$.

b) Present the confusion matrix and the apparent error rate of misclassification (APER) for your classification rule of part a).

c) Suppose we are interested in a new graduate whose score for math and literature are 17 and 16, respectively. Would you classify him as a graduate from HKU or from HKUST based on your classfication rule in part a)?

d) In a business firm, employees that are graduates of HKU are three times as many as employees that are graduates from HKUST. One employee is known to be a graduate from one of the two universities. He took the test and scored also 17 for math and 16 for literature. Would you classify him as a graduate from HKU or from HKUST? Assume still the costs of the two types of misclassifications are same.

e) Plot the six observations and your classfication rules in both part a) and part d). Also indicate in your plot the new observation who scored 17 and 16 for math and literature tests.