Exam I (In class portion 10%)

Show your detailed work for full credit

1. (1 pt) The degrees of morning temperature (X_1) and afternoon temperature (X_2) in Celsius for last three days are organized into the data matrix $\mathbf{X}$. $\mathbf{X} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$. Compute Cov ($\mathbf{X}$), the sample variance-covariance matrix of $\mathbf{X}$.

$S_{xy} = \frac{1}{n-1} \Sigma (x_i - \bar{x})(y_i - \bar{y})$ $\quad n = 3. \quad \bar{x} = 2 \quad \bar{y} = 3$

$S_{xx} = \frac{1}{2} [(4-2)^2 + (-1-2)^2 + (3-2)^2]$

$= \frac{14}{2} = 7$

$S_{yy} = \frac{1}{2} [(1-3)^2 + (3-3)^2 + (5-3)^2]$

$= 4$

$S_{xy} = \frac{1}{2} [(4-2)(1-3) + (-1-2)(3-3) + (3-2)(5-3)]$

$= \frac{1}{2} [-4 + 0 + 2]$

$= -1$

$\text{cov} = \begin{pmatrix} 7 & -1 \\ -1 & 4 \end{pmatrix}$

2. Suppose the sample correlation matrix of a bivariate data is $\mathbf{R} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$. $r > 0$

a. (1 pt) Find the eigenvalues of $\mathbf{R}$.

$$\det(R - \lambda I) = \begin{vmatrix} 1-\lambda & r \\ r & 1-\lambda \end{vmatrix} = (1 - \lambda + r)(1 - \lambda - r) = 0$$

$$(1-\lambda+r) = 0 \qquad 1-\lambda-r = 0$$

$$\lambda_1 = 1+r > \lambda_2 = 1-r$$

b. (1 pt) Find the normalized eigenvectors of $\mathbf{R}$.

$$\lambda_1 = 1+r. \quad (R-\lambda I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} -r & r \\ +r & -r \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-x_1 + x_2 = 0 \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \underset{\text{normalize}}{\Rightarrow} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\lambda_2 = 1-\lambda \quad (R-\lambda I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} r & r \\ r & r \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 + x_2 = 0 \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \underset{\text{normalize}}{\Rightarrow} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

c. (1 pt) What are the two principal components?

$$Y_1 = \frac{1}{\sqrt{2}}x_1 + \frac{1}{\sqrt{2}}x_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a^T\vec{x} \qquad \vec{a}^T = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$Y_2 = \frac{1}{\sqrt{2}}x_1 - \frac{1}{\sqrt{2}}x_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = b^T\vec{x} \qquad b^T = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

d. (1 pt) Calculate the sample variances of the two principal components.

$$Var(Y_1) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) cov(x) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}}(1, 1)\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$= 1+r$$

$$Var(Y_2) = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) cov(x) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2}(1, -1)\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$
$$= 1-r.$$

3. (3 pts) Explain how to find the first principal component Y_1 and derive that $Var(Y_1) = \lambda_1$, where λ_1 is the largest eigenvalue of the variance-covariance matrix.

Suppose R is covariance matrix of multivariable X. λ_1 is the largest eigenvalue of R and $(a_{11}, a_{12}, \cdots a_{1g})^T$ is a standardized eigenvector correspounding to λ_1. then the first principal component $Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1g}X_g$

$$= \vec{a}_1^T X$$

$$Var(Y_1) = \vec{a}_1^T \underline{cov(X)\vec{a}_1} \qquad cov(X)\vec{a}_1 = \lambda\vec{a}_1$$

$$= \vec{a}_1^T \lambda_1 \vec{a}_1$$

$$= \lambda_1 \vec{a}_1^T \vec{a}_1$$

$$= \lambda_1$$

4. (2 pts) Let $X_1, X_2, \cdots, X_q$ are q multivariate variables with a sample-variance covaraince matrix of $\mathbf{S}$. The principal components of the variables, $Y_j,\ j = 1, 2, \cdots, q$, are the linear combinations of the X_i's, where $i = 1, 2, \cdots, q$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_j \geq \cdots \geq \lambda_q$ are eigenvalues and $\mathbf{a}_j = (a_{j1}, a_{j2}, \cdots, a_{jq})^T$ are correspounding eigenvectors.

Show that the correlation of variable x_i with component y_j is $r_{x_i;y_j} = \frac{a_{ji}\sqrt{\lambda_j}}{s_{x_i}}$.

$$r_{x_i, y_i} = \frac{cov(X_i, y_j)}{S_{x_i} S_{y_j}} = \frac{cov(X_i, a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jg}X_g)}{S_{x_i} S_{y_j}}$$

$$= \frac{\overset{(i)}{(0, \cdots 0, 1, 0 \cdots 0)}\, cov(X) \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jg} \end{pmatrix}}{S_{x_i} S_{y_j}}$$

$$= \frac{(0, \cdots 0, 1, 0 \cdots 0)\, \lambda_j \vec{a}_j}{S_{x_i} S_{y_j}} \qquad \vec{a}_j = \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{ji} \\ \vdots \\ a_{jg} \end{pmatrix}$$

$$= \frac{\lambda_j a_{ji}}{S_{x_i}\sqrt{\lambda_j}} = \frac{\sqrt{\lambda_j}\, a_{ji}}{S_{x_i}}$$