Jared Yu 914640019

HW 7 STA 135

1. Suppose the population mean and covariance matrix of $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ are

$$\vec{\mu} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{bmatrix}.$$

   a. Determine the first and second principal components Y_1 and Y_2, and find their variance respectively.

$$\mathbf{\Sigma} - \lambda I = 0$$

$$\begin{bmatrix} 2-\lambda & \sqrt{3} \\ \sqrt{3} & 4-\lambda \end{bmatrix} = 0$$

$$(2-\lambda)(4-\lambda) - 3 = 0$$

$$\lambda^2 - 6\lambda + 5 = 0$$

$$(\lambda - 5)(\lambda - 1) = 0$$

$$\lambda_1 = 5, \lambda_2 = 1$$

In the case of $\lambda_1 = 5$:

$$\begin{bmatrix} 2-5 & \sqrt{3} \\ \sqrt{3} & 4-5 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{bmatrix} -3 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$R_1 + \sqrt{3}R_2 \to R_2 \begin{bmatrix} -3 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$-\frac{1}{3}R_1 \to R_1 \begin{bmatrix} 1 & -\frac{\sqrt{3}}{3} \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$x - \frac{\sqrt{3}}{3}y = 0$$

$$x = \frac{\sqrt{3}}{3}y$$

$$e_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}$$

In the case of $\lambda_2 = 1$:

$$\begin{bmatrix} 2-1 & \sqrt{3} \\ \sqrt{3} & 4-1 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$R_1 - \frac{1}{\sqrt{3}}R_2 \to R_2 \begin{bmatrix} 1 & \sqrt{3} \\ 0 & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$x + \sqrt{3}y = 0$$

$$x = -\sqrt{3}y$$

$$e_2 = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\boxed{Y_1 = e_1' \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \frac{1}{2}X_1 + \frac{\sqrt{3}}{2}X_2 = 0.5X_1 + 0.866X_2}$$

$$Var(Y_1) = Var\left(e_1' \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) = e_1' Var\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) e_1 = e_1' \mathbf{\Sigma} e_1 = e_1' \lambda_1 e_1 = \lambda_1 = \boxed{5}$$

$$\boxed{Y_2 = e_2' \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = -\frac{\sqrt{3}}{2}X_1 + \frac{1}{2}X_2 = -0.866X_1 + 0.5X_2}$$

$$Var(Y_2) = Var\left(e_2' \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) = e_2' Var\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) e_2 = e_2' \mathbf{\Sigma} e_2 = e_2' \lambda_2 e_2 = \lambda_2 = \boxed{1}$$

b. Determine the proportion of total variance due to the Y_1.

$$(Proportion\ of\ total\ population\ variance\ due\ to\ 1st\ principal\ component) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

$$= \frac{5}{5+1} = \boxed{\frac{5}{6} = 0.8\overline{3}}$$

c. Compare the contributions of X_1 and X_2 to the determination of Y_1 based on loadings and correlations, respectively.

In the case of Y_1, the loading of X_2 is much larger at 0.866 in comparison to the loading of X_1 at 0.5. So, the loading on X_2 dominates the loading on X_1.

$$Corr(Y_1, X_1) = v_{11}\sqrt{\frac{\lambda_1}{\sigma_{11}}} = \frac{1}{2}\sqrt{\frac{5}{2}} = 0.791$$

$$Corr(Y_1, X_2) = v_{12}\sqrt{\frac{\lambda_1}{\sigma_{22}}} = \frac{\sqrt{3}}{2}\sqrt{\frac{5}{4}} = \frac{\sqrt{15}}{4} = 0.968$$

In the case of correlations, the contribution of X_2 is 0.968, while the contribution of X_1 is 0.791. The correlation does not show the dominance of X_1 over X_1 as was seen in the case of just the loadings. The correlation shows again that X_2 is closer to being important to Y_1.

2. Suppose that the random vector $\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ has the following population mean and covariance matrix:

$$\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

a. Determine the first two principal components Y_1 and Y_2 and the proportion of total variance due to them. For $i = 1,2$, compare the contributions of X_1, X_2, and X_3 to the determination of Y_i based on loadings and correlations, respectively.

$$\boldsymbol{\Sigma} - \lambda \boldsymbol{I} = 0$$

$$\begin{vmatrix} 3-\lambda & 0 & 0 \\ 0 & 3-\lambda & 1 \\ 0 & 1 & 3-\lambda \end{vmatrix} = 0$$

$$(3-\lambda)[(3-\lambda)^2 - 1] = 0$$

$$(3-\lambda)[(4-\lambda)(2-\lambda)] = 0$$

$$\lambda_1 = 4\ \lambda_2 = 3, \lambda_3 = 2$$

$$\lambda_1 = 4 \rightarrow \begin{pmatrix} 3-4 & 0 & 0 \\ 0 & 3-4 & 1 \\ 0 & 1 & 3-4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$R_2 + R_3 \rightarrow R_3 \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$-x = 0$$

$$x = 0$$

$$-y + z = 0$$

$$z = y$$

$$e_1 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\lambda_2 = 3 \rightarrow \begin{pmatrix} 3-3 & 0 & 0 \\ 0 & 3-3 & 1 \\ 0 & 1 & 3-3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$x \in \mathbb{R}$$

$$y = 0$$

$$z = 0$$

$$e_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\lambda_3 = 2 \rightarrow \begin{pmatrix} 3-2 & 0 & 0 \\ 0 & 3-2 & 1 \\ 0 & 1 & 3-2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$R_2 - R_3 \rightarrow R_3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$x = 0$$

$$y + z = 0$$

$$y = -z$$

$$e_3 = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$(Proportion\ of\ total\ population\ variance\ due\ to\ 1st\ and\ 2nd\ principal\ component)$$

$$= \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{4+3}{4+3+2} = \boxed{\frac{7}{9} = 0.\overline{7}}$$

$$Y_1 = e_1' \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = 0X_1 + \frac{1}{\sqrt{2}}X_2 + \frac{1}{\sqrt{2}}X_3 = 0.707X_2 + 0.707X_3$$

$$Y_2 = e_2' \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = 1X_1 + 0X_2 + 0X_3 = X_1$$

$$Corr(Y_1, X_1) = v_{11}\sqrt{\frac{\lambda_1}{\sigma_{11}}} = 0$$

$$Corr(Y_1, X_2) = v_{12}\sqrt{\frac{\lambda_1}{\sigma_{22}}} = \frac{1}{\sqrt{2}}\sqrt{\frac{4}{3}} = \frac{2}{\sqrt{6}} = \frac{\sqrt{6}}{3} = 0.816$$

$$Corr(Y_1, X_3) = v_{13}\sqrt{\frac{\lambda_1}{\sigma_{33}}} = \frac{1}{\sqrt{2}}\sqrt{\frac{4}{3}} = \cdots = \frac{\sqrt{6}}{3} = 0.816$$

$$Corr(Y_2, X_1) = v_{21}\sqrt{\frac{\lambda_2}{\sigma_{11}}} = 1\sqrt{\frac{3}{3}} = 1$$

$$Corr(Y_2, X_2) = v_{22}\sqrt{\frac{\lambda_2}{\sigma_{22}}} = \cdots = 0$$

$$Corr(Y_2, X_3) = v_{23}\sqrt{\frac{\lambda_2}{\sigma_{33}}} = \cdots = 0$$

In Y_1, the contribution of X_2 and X_3 are equal at 0.707, but the contribution of X_1 is 0 in terms of loadings. In Y_2, the contribution of X_1 is 1, while the contribution of X_2 and X_3 is 0 in terms of loadings. The correlations in Y_1 follow a similar pattern, where only X_2 and X_3 contribute with an equal correlation of 0.816 showing that they're equally important. The correlation for Y_2 also follows a similar pattern where X_1 dominates in terms of importance while the other two variables have no correlation.

b. Let Z_1, Z_2, and Z_3 be the standardized variables of X_1, X_2, and X_3, respectively. Find the first two principal components W_1 and W_2 of (Z_1, Z_2, Z_3). For $i = 1,2$, compare the contributions of Z_1, Z_2, and Z_3 to the determination of W_i based on loadings and correlations, respectively.

First, let

$$\boldsymbol{V}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & & \\ & \frac{1}{\sqrt{\sigma_{22}}} & \\ & & \frac{1}{\sqrt{\sigma_{33}}} \end{bmatrix},$$

then,

$$\boldsymbol{\rho} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\Sigma}\left(\boldsymbol{V}^{-\frac{1}{2}}\right)^T$$

$$\begin{bmatrix} \frac{\sqrt{3}}{3} & & \\ & \frac{\sqrt{3}}{3} & \\ & & \frac{\sqrt{3}}{3} \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{3}}{3} & & \\ & \frac{\sqrt{3}}{3} & \\ & & \frac{\sqrt{3}}{3} \end{bmatrix} = \frac{\sqrt{3}}{3}\boldsymbol{I}\boldsymbol{\Sigma}\left(\frac{\sqrt{3}}{3}\boldsymbol{I}\right)^T = \frac{1}{3}\boldsymbol{\Sigma}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{3} \\ 0 & \frac{1}{3} & 1 \end{bmatrix}.$$

Then solving for the eigenvalues and eigenvectors of $\boldsymbol{\rho}$:

$$\boldsymbol{\rho} - \lambda\boldsymbol{I} = 0$$

$$\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & \frac{1}{3} \\ 0 & \frac{1}{3} & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)\left[\left(\frac{4}{3}-\lambda\right)\left(\frac{2}{3}-\lambda\right)\right] = 0$$

$$\lambda_1 = \frac{4}{3}, \lambda_2 = 1, \lambda_3 = \frac{2}{3}$$

$$\lambda_1 = \frac{4}{3} \rightarrow \begin{pmatrix} 1-\frac{4}{3} & 0 & 0 \\ 0 & 1-\frac{4}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & 1-\frac{4}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & -\frac{1}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$R_2 + R_3 \to R_3, \begin{pmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$-3R_1 \to R_1, -3R_2 \to R_2, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$x = 0$$

$$y - z = 0$$

$$y = z$$

$$e_1 = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$$

$$\lambda_2 = 1 \to \begin{pmatrix} 1-1 & 0 & 0 \\ 0 & 1-1 & \frac{1}{3} \\ 0 & \frac{1}{3} & 1-1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$3R_2 \to R_2, 3R_3 \to R_3, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$x \in \mathbb{R}$$

$$z = 0$$

$$y = 0$$

$$e_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$W_1 = e_1' \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = 0\left(\frac{X_1 - \mu_1}{\sqrt{3}}\right) + \frac{\sqrt{2}}{2}\left(\frac{X_2 - \mu_2}{\sqrt{3}}\right) + \frac{\sqrt{2}}{2}\left(\frac{X_3 - \mu_3}{\sqrt{3}}\right)$$

$$= \boxed{\frac{\sqrt{2}}{2}\left(\frac{X_2 - \mu_2}{\sqrt{3}}\right) + \frac{\sqrt{2}}{2}\left(\frac{X_3 - \mu_3}{\sqrt{3}}\right)}$$

$$W_2 = e_2' \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = 1\left(\frac{X_1 - \mu_1}{\sqrt{3}}\right) + 0\left(\frac{X_2 - \mu_2}{\sqrt{3}}\right) + 0\left(\frac{X_3 - \mu_3}{\sqrt{3}}\right) = \boxed{\left(\frac{X_1 - \mu_1}{\sqrt{3}}\right)}$$

$$Corr(W_1, Z_1) = e_{11}\sqrt{\lambda_1} = \cdots = 0$$

$$Corr(W_1, Z_2) = e_{12}\sqrt{\lambda_1} = \frac{\sqrt{2}}{2}\sqrt{\frac{4}{3}} = \frac{\sqrt{6}}{3} = 0.816$$

$$Corr(W_1, Z_3) = e_{13}\sqrt{\lambda_1} = \cdots = 0.816$$

$$Corr(W_2, Z_1) = e_{21}\sqrt{\lambda_2} = 1\sqrt{1} = 1$$

$$Corr(W_2, Z_2) = e_{22}\sqrt{\lambda_2} = \cdots = 0$$

$$Corr(W_2, Z_3) = e_{23}\sqrt{\lambda_2} = \cdots = 0$$

The contributions of Z_2 and Z_3 to W_1 are the same at 0.707 while Z_1 has a loading of 0. The correlation for W_1 is similar where Z_2 and Z_3 have a correlation of 0.816 while Z_1 has a correlation of 0. So Z_2 and Z_3 are equally important for W_1. The contributions of Z_2 and Z_3 in W_2 have a loading of 0 while Z_1 has a loading of 1. The correlation is similar in W_2 where both Z_2 and Z_3 have a correlation of 0 while Z_1 has a correlation of 1 with W_2. For W_2, only Z_1 is important for the determination.

3. Suppose a sample $\vec{x}_1, \cdots, \vec{x}_n$ has the sample mean $\bar{\vec{x}}$ and sample covariance $\boldsymbol{S}$. Let $\hat{y}_i$ be the i-th sample principal component, where $i = 1, \cdots, p$. Then we can transform the data matrix

$$\boldsymbol{X} = \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \cdots \\ \vec{x}_n^T \end{bmatrix}$$

into the data matrix of the first $r(< p)$ sample principal components

$$\widehat{\boldsymbol{Y}} = \begin{bmatrix} \hat{y}_{11} & \hat{y}_{12} & \cdots & \hat{y}_{1r} \\ \hat{y}_{21} & \hat{y}_{22} & \cdots & \hat{y}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{n1} & \hat{y}_{n2} & \cdots & \hat{y}_{nr} \end{bmatrix}.$$

Please find the relationship between $\boldsymbol{X}$ and $\widehat{\boldsymbol{Y}}$ with the spectral decomposition of $\boldsymbol{S}$.

Let $\boldsymbol{S} = \hat{\lambda}_1 \vec{u}_1 \vec{u}_1^\top + \cdots + \hat{\lambda}_p \vec{u}_p \vec{u}_p^\top = \boldsymbol{U}\widehat{\boldsymbol{\Lambda}}\boldsymbol{U}^\top$ be the spectral decomposition of $\boldsymbol{S}$. Here $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$, $\boldsymbol{U} = [\vec{u}_1 \quad \vec{u}_2 \quad \cdots \quad \vec{u}_p]$, and

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} \hat{\lambda}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_k \end{bmatrix}.$$

The i-th observation of all sample principal components is

$$\hat{\vec{y}}_i = \begin{bmatrix} \hat{y}_{i1} \\ \vdots \\ \hat{y}_{ip} \end{bmatrix} = \begin{bmatrix} u_{11}x_{i1} + u_{12}x_{i2} + \cdots + u_{1p}x_{ip} \\ \vdots \\ u_{p1}x_{i1} + u_{p2}x_{i2} + \cdots + u_{pp}x_{ip} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1p} \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pp} \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} = \boldsymbol{U}^\top \vec{x}_i.$$

Then it follows that the unreduced version of the data matrix of sample principal components is

$$\widehat{\boldsymbol{Y}}_{unreduced} = \underbrace{\begin{bmatrix} \hat{\vec{y}}_1^\top \\ \hat{\vec{y}}_2^\top \\ \vdots \\ \hat{\vec{y}}_n^\top \end{bmatrix}}_{n\times p} = \begin{bmatrix} \hat{\vec{x}}_1^\top \boldsymbol{U} \\ \hat{\vec{x}}_2^\top \boldsymbol{U} \\ \vdots \\ \hat{\vec{x}}_n^\top \boldsymbol{U} \end{bmatrix} = \begin{bmatrix} \hat{\vec{x}}_1^\top \\ \hat{\vec{x}}_2^\top \\ \vdots \\ \hat{\vec{x}}_n^\top \end{bmatrix} \boldsymbol{U} = \underbrace{\boldsymbol{X}}_{n\times p} \underbrace{\boldsymbol{U}}_{p\times p}.$$

Let $\widetilde{\boldsymbol{U}}$ be a reduced version of $\boldsymbol{U}$ and $\widetilde{\boldsymbol{I}}$ is a $p \times r$ identity matrix $\boldsymbol{I}$ where all the rows and only the first r columns are retained such that,

$$\underbrace{\widetilde{\boldsymbol{U}}}_{p\times r} = \underbrace{\boldsymbol{U}}_{p\times p} \underbrace{\widetilde{\boldsymbol{I}}}_{p\times r}.$$

So, it could be said that,

$$\boxed{\underbrace{\widehat{\boldsymbol{Y}}}_{n\times r} = \underbrace{\boldsymbol{X}}_{n\times p} \underbrace{\widetilde{\boldsymbol{U}}}_{p\times r}.}$$

4. You are given a sample $\vec{x}_1, \cdots, \vec{x}_6$ from a 2-dimension population. Moreover, the sample principal components are

$$\hat{y}_{i1} = \frac{\sqrt{2}}{2}x_{i1} + \frac{\sqrt{2}}{2}x_{i2}, \hat{y}_{i2} = \frac{\sqrt{2}}{2}x_{i1} - \frac{\sqrt{2}}{2}x_{i2}, i = 1, \cdots, 6.$$

Assume the sample variances of $(\hat{y}_{i1})_{i=1}^{6}$ and $(\hat{y}_{i2})_{i=1}^{6}$ are 3 and 2, respectively. Find the sample covariance matrix of the original data matrix $\boldsymbol{X}$.

From the equations for $\hat{y}_{i1}$ and $\hat{y}_{i2}$ it can be determined that,

$$e_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$$

and

$$e_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Then from the sample variances it can also be determined that $\lambda_1 = 3$ and $\lambda_2 = 2$. This follows directly from problem 1, part a. Given that the eigenvectors and eigenvalues of the sample covariance matrix are known, then the original sample covariance matrix, $\boldsymbol{S}$, can be found through the following formula,

$$\boldsymbol{S} = \boldsymbol{U\Lambda U}^{-1}$$

where $\boldsymbol{U} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$ and $\boldsymbol{\Lambda} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. Then it follows that,

$$\boldsymbol{S} = \boldsymbol{U\Lambda U}^{-1}$$

$$= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} = \cdots$$

Note: $\boldsymbol{U}^{-1} = \boldsymbol{U}$ since it is orthogonal.

$$\cdots = \begin{bmatrix} \frac{3\sqrt{2}}{2} & \sqrt{2} \\ \frac{3\sqrt{2}}{2} & -\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} = \boxed{\begin{bmatrix} \frac{5}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{5}{2} \end{bmatrix}}$$

5. Consider a sample of $\vec{x}_{11}, \cdots, \vec{x}_{1n_1}$ of size $n_1 = 10$ from population 1 (corresponding to class π_1) and a sample of $\vec{x}_{11}, \cdots, \vec{x}_{1n_2}$ of size $n_2 = 10$ from population 2 (corresponding to π_2). The summary statistics for these two samples are

$$\bar{\vec{x}}_1 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \boldsymbol{S}_1 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix},$$

$$\bar{\vec{x}}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \boldsymbol{S}_2 = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}.$$

For some $\vec{x}_0 = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}$, derive the following classifiers:

1. Classifier 1: Fisher's rule based on $\vec{x}_0$, and give the D^2 distance denoted as D_1^2;

$$D_1^2 = \left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)^\top \boldsymbol{S}_{pooled}^{-1} \left(\vec{x}_0 - \frac{1}{2}\left(\bar{\vec{x}}_1 + \bar{\vec{x}}_2\right)\right),$$

where $\vec{x}_0$ is assigned to π_1 if $D_1^2 \geq 0$, and π_2 otherwise. Then finding

$$\boldsymbol{S}_{pooled} = \frac{n_1 - 1}{n_1 + n_2 - 2}\boldsymbol{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\boldsymbol{S}_2 = \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} + \begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

and

$$\boldsymbol{S}_{pooled}^{-1} = \frac{1}{8}\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{bmatrix}.$$

$$D_1^2 = \left(\begin{bmatrix} 6 \\ -3 \end{bmatrix}\right)^\top \begin{bmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{bmatrix} \left(\begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} - \frac{1}{2}\left(\begin{bmatrix} 6 \\ 3 \end{bmatrix}\right)\right) = \left(\begin{bmatrix} 6 \\ -3 \end{bmatrix}\right)^\top \begin{bmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{bmatrix} \left(\begin{bmatrix} x_{01} - 3 \\ x_{02} - \frac{3}{2} \end{bmatrix}\right)$$

$$= \begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} x_{01} - 3 \\ x_{02} - \frac{3}{2} \end{bmatrix} = \frac{21}{8}(x_{01} - 3) - \frac{15}{8}\left(x_{02} - \frac{3}{2}\right) = \frac{21}{8}x_{01} - \frac{63}{8} - \frac{15}{8}x_{02} + \frac{45}{16}$$

$$= \frac{21}{8}x_{01} - \frac{15}{8}x_{02} - \frac{81}{16} = \boxed{2.63x_{01} - 1.88x_{02} - 5.06}$$

2. Classifier 2: Fisher's rule based on x_{01}, and give the D^2 distance denoted as D_2^2;

Let $\boldsymbol{C}^\top = [1 \quad 0]$, then

$$D_2^2 = \left(\boldsymbol{C}^\top \bar{\vec{x}}_1 - \boldsymbol{C}^\top \bar{\vec{x}}_2\right)^\top \left(\boldsymbol{C}^\top \boldsymbol{S}_{pooled} \boldsymbol{C}\right)^{-1} \left(\boldsymbol{C}^\top \vec{x}_0 - \frac{1}{2}\boldsymbol{C}^\top\left(\bar{\vec{x}}_1 + \bar{\vec{x}}_2\right)\right)$$

$$= (6)^{\top}\left(\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}3 & 1\\1 & 3\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right)^{-1}(x_{01}-3) = 2(x_{01}-3) = \boxed{2x_{01}-6}$$

where x_{01} is assigned to π_1 if $D_2^2 \geq 0$, and π_2 otherwise.

3. Classifier 3: Fisher's rule based on x_{02}, and give the D^2 distance denoted as D_3^2;
   Let $\boldsymbol{C}^{\top} = [0 \quad 1]$, then

$$D_3^2 = \left(\boldsymbol{C}^{\top}\bar{\vec{x}}_1 - \boldsymbol{C}^{\top}\bar{\vec{x}}_2\right)^{\top}\left(\boldsymbol{C}^{\top}\boldsymbol{S}_{pooled}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}^{\top}\vec{x}_0 - \frac{1}{2}\boldsymbol{C}^{\top}\left(\bar{\vec{x}}_1 + \bar{\vec{x}}_2\right)\right)$$

$$= (-3)^{\top}\left(\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}3 & 1\\1 & 3\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\right)^{-1}\left(x_{02}-\frac{3}{2}\right) = -1\left(x_{02}-\frac{3}{2}\right) = \boxed{-x_{02}+1.5}$$

where x_{02} is assigned to π_1 if $D_3^2 \geq 0$, and π_2 otherwise.

4. Classifier 4: Fisher's rule based on $\vec{x}_0^T\left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)$, and give the D^2 distance denoted as D_4^2;

   Let $\boldsymbol{C}^{\top} = \left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)^{\top}$, then

$$D_4^2 = \left(\boldsymbol{C}^{\top}\bar{\vec{x}}_1 - \boldsymbol{C}^{\top}\bar{\vec{x}}_2\right)^{\top}\left(\boldsymbol{C}^{\top}\boldsymbol{S}_{pooled}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}^{\top}\vec{x}_0 - \frac{1}{2}\boldsymbol{C}^{\top}\left(\bar{\vec{x}}_1 + \bar{\vec{x}}_2\right)\right)$$

$$= \left(\begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}6\\0\end{bmatrix} - \begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}0\\3\end{bmatrix}\right)^{\top}\left(\begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}3 & 1\\1 & 3\end{bmatrix}\begin{bmatrix}6\\-3\end{bmatrix}\right)^{-1}\left(\begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}x_{01}\\x_{02}\end{bmatrix} - \frac{1}{2}\begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}6\\3\end{bmatrix}\right)$$

$$= (45)^{\top}([81])^{-1}\left([6x_{01}-3x_{02}] - \frac{27}{2}\right) = \frac{5}{9}\left(6x_{01}-3x_{02}-\frac{27}{2}\right) = \frac{10}{3}x_{01} - \frac{5}{3}x_{02} - \frac{15}{2}$$
$$= \boxed{3.33x_{01} - 1.67x_{02} - 7.5}$$

where $\vec{x}_0^T\left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)$ is assigned to π_1 if $D_4^2 \geq 0$, and π_2 otherwise.

5. Classifier 5: Fisher's rule based on $\vec{x}_0^T\boldsymbol{S}_{pooled}^{-1}\left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)$, and give the D^2 distance denoted as D_5^2;

   Let $\boldsymbol{C}^{\top} = \left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)^{\top}\boldsymbol{S}_{pooled}^{-1}$, then

$$D_5^2 = \left(\boldsymbol{C}^{\top}\bar{\vec{x}}_1 - \boldsymbol{C}^{\top}\bar{\vec{x}}_2\right)^{\top}\left(\boldsymbol{C}^{\top}\boldsymbol{S}_{pooled}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}^{\top}\vec{x}_0 - \frac{1}{2}\boldsymbol{C}^{\top}\left(\bar{\vec{x}}_1 + \bar{\vec{x}}_2\right)\right)$$

$$Note: \boldsymbol{C}^{\top} = \left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)^{\top}\boldsymbol{S}_{pooled}^{-1} = \begin{bmatrix}6\\-3\end{bmatrix}^{\top}\begin{bmatrix}\frac{3}{8} & -\frac{1}{8}\\-\frac{1}{8} & \frac{3}{8}\end{bmatrix} = \begin{bmatrix}\frac{21}{8} & -\frac{15}{8}\end{bmatrix}$$

$$= \left(\begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} \right)^{\top} \left(\begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{21}{8} \\ -\frac{15}{8} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} - \frac{1}{2} \begin{bmatrix} \frac{21}{8} & -\frac{15}{8} \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} \right)$$

$$= \left(\frac{81}{8}\right)^{\top} \left(\frac{64}{1368}\right) \left(\frac{21}{8}x_{01} - \frac{15}{8}x_{02} - \frac{81}{16}\right) = \left(\frac{9}{19}\right) \left(\frac{21}{8}x_{01} - \frac{15}{8}x_{02} - \frac{81}{16}\right)$$

$$= \boxed{1.24x_{01} - 0.89x_{02} - 2.40}$$

where $\vec{x}_0^T\left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)$ is assigned to π_1 if $D_5^2 \geq 0$, and π_2 otherwise.

6. Compare the above D^2 distances.

$$D_1^2 = 2.63x_{01} - 1.88x_{02} - 5.06$$
$$D_2^2 = 2x_{01} - 6$$
$$D_3^2 = -x_{02} + 1.5$$
$$D_4^2 = 3.33x_{01} - 1.67x_{02} - 7.5$$
$$D_5^2 = 1.24x_{01} - 0.89x_{02} - 2.40$$

The above distances will classify to π_1 if they are ≥ 0, and π_2 otherwise.

6. Suppose we have n_1 p-variate observations from π_1 and n_2 p-variate observations from π_2. The respective data matrices are

$$\boldsymbol{X}_1 = \begin{bmatrix} \vec{x}_{11}^T \\ \vec{x}_{12}^T \\ \cdots \\ \vec{x}_{1n_1}^T \end{bmatrix}, \boldsymbol{X}_2 = \begin{bmatrix} \vec{x}_{21}^T \\ \vec{x}_{22}^T \\ \cdots \\ \vec{x}_{2n_2}^T \end{bmatrix}.$$

Suppose $\bar{\vec{x}}$ is the overall sample mean of these two samples. Consider the data matrices

$$\boldsymbol{Z}_1 = \begin{bmatrix} \vec{z}_{11}^T \\ \vec{z}_{12}^T \\ \cdots \\ \vec{z}_{1n_1}^T \end{bmatrix}, \boldsymbol{Z}_2 = \begin{bmatrix} \vec{z}_{21}^T \\ \vec{z}_{22}^T \\ \cdots \\ \vec{z}_{2n_2}^T \end{bmatrix},$$

where $\vec{z}_{lj} = \vec{x}_{lj} - \bar{\vec{x}}, l = 1,2, j = 1, \cdots, n_l$. Show that by Fisher's linear discriminant, $\vec{x}_0$ is allocated to the first population based on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ if and only if $\vec{x}_0 - \bar{\vec{x}}$ is allocated to the first population based on $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$.

$\vec{x}_0 - \bar{\vec{x}}$ is allocated to the first population based on $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ iff

$$\left(\bar{\vec{z}}_1 - \bar{\vec{z}}_2\right)^\top \boldsymbol{S}_{pooled}^{-1} \left(\vec{z} - \frac{1}{2}\left(\bar{\vec{z}}_1 + \bar{\vec{z}}_2\right)\right) \geq 0.$$

iff

$$\left(\left(\vec{x}_1 - \bar{\vec{x}}\right) - \left(\vec{x}_2 - \bar{\vec{x}}\right)\right)^\top \boldsymbol{S}_{pooled}^{-1} \left(\left(\vec{x}_0 - \bar{\vec{x}}\right) - \frac{1}{2}\left(\left(\vec{x}_1 - \bar{\vec{x}}\right) + \left(\vec{x}_2 - \bar{\vec{x}}\right)\right)\right) \geq 0$$

iff

$$= (\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1} \left(\left(\vec{x}_0 - \bar{\vec{x}}\right) - \frac{1}{2}\left(\vec{x}_1 + \vec{x}_2 - 2\bar{\vec{x}}\right)\right) \geq 0$$

iff

$$= (\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1} \left(\vec{x}_0 - \frac{1}{2}(\vec{x}_1 + \vec{x}_2)\right) \geq 0.$$

Then $\vec{x}_0$ is allocated to the first population based on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

7. Consider three independent samples from three classes:

π_1: distribution $N_p(\vec{\mu}_1, \boldsymbol{\Sigma})$, sample size n_1, sample mean $\bar{\vec{x}}_1$, sample covariance $\boldsymbol{S}_1$;

π_2: distribution $N_p(\vec{\mu}_2, \boldsymbol{\Sigma})$, sample size n_2, sample mean $\bar{\vec{x}}_2$, sample covariance $\boldsymbol{S}_2$;

π_3: distribution $N_p(\vec{\mu}_3, \mathbf{\Sigma})$, sample size n_3, sample mean $\bar{\vec{x}}_3$, sample covariance $\boldsymbol{S}_3$.

Assume that these three populations have equal prior probabilities. For a new observation $\vec{x}_0$, we aim to classify it to one of three classes with pairwise linear discriminant analyses based on Fisher's rule. Given the three population covariance matrices are assumed to be the same, in all linear discriminant analyses we use

$$\boldsymbol{S}_{pooled} = \frac{1}{n_1 + n_2 + n_3 - 3}\big((n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2 + (n_3 - 1)\boldsymbol{S}_3\big).$$

Supposed that in the comparison between π_1 and π_2, $\vec{x}_0$ is allocated to π_2, while in the comparison between π_2 and π_3, $\vec{x}_0$ is allocated to π_3. Show that in the comparison between π_1 and π_3, $\vec{x}_0$ is allocated to π_3.

If

$$(\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1}\left(\vec{x}_0 - \frac{1}{2}(\vec{x}_1 + \vec{x}_2)\right) < 0$$

$$(\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_0 < \frac{1}{2}(\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_1 + \vec{x}_2)$$

(using ECM rule for two normal populations p.558 in text)

then $\vec{x}_0$ is allocated to π_2.

If

$$(\vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}\left(\vec{x}_0 - \frac{1}{2}(\vec{x}_2 + \vec{x}_3)\right) < 0$$

$$(\vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_0 < \frac{1}{2}(\vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_2 + \vec{x}_3)$$

then $\vec{x}_0$ is allocated to π_3.

Then it must be shown that,

$$(\vec{x}_1 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}\left(\vec{x}_0 - \frac{1}{2}(\vec{x}_1 + \vec{x}_3)\right) < 0$$

$$(\vec{x}_1 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_0 < \frac{1}{2}(\vec{x}_1 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_1 + \vec{x}_3)$$

in order to allocate $\vec{x}_0$ is allocated to π_3.

Starting from the L.H.S.,

$$(\vec{x}_1 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1} \vec{x}_0 = (\vec{x}_1 - \vec{x}_2 + \vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1} \vec{x}_0$$

$$= (\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1} \vec{x}_0 + (\vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1} \vec{x}_0$$

$$< \frac{1}{2}(\vec{x}_1 - \vec{x}_2)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_1 + \vec{x}_2) + \frac{1}{2}(\vec{x}_2 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_2 + \vec{x}_3)$$

$$= \frac{1}{2}\vec{x}_1^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 + \frac{1}{2}\vec{x}_1^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_2 - \frac{1}{2}\vec{x}_2^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 - \frac{1}{2}\vec{x}_2^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_2 + \frac{1}{2}\vec{x}_2^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_2^\top$$
$$+ \frac{1}{2}\vec{x}_2^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_3^\top - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_2^\top\vec{x}_2 - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_3^\top$$

$$= \frac{1}{2}\vec{x}_1^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_3$$

$$= \frac{1}{2}\vec{x}_1^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 + \frac{1}{2}\vec{x}_1^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_3 - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_1 - \frac{1}{2}\vec{x}_3^\top \boldsymbol{S}_{pooled}^{-1}\vec{x}_3$$

$$= \frac{1}{2}(\vec{x}_1 - \vec{x}_3)^\top \boldsymbol{S}_{pooled}^{-1}(\vec{x}_1 + \vec{x}_3)$$

which is the R.H.S.