1. Fit a multivariate straight line model using the given data and prove

$$Y^T Y = \hat{Y}^T \hat{Y} + \hat{\varepsilon}^T \hat{\varepsilon}$$

Z^T=(2, 3, 4, 5, 6); $Y_1^T = (2, 6, 8, 9, 12)$; $Y_1^T = (-2, -2, 4, 6, 4)$ (20)

Design matrix$-$1 $marks$

$\hat{\beta}$- formula – 1 marks

$\hat{\beta}_1$- 3 marks

$\hat{\beta}_2$- 3 marks $\begin{bmatrix} -1.8 & -6 \\ 2.3 & 2 \end{bmatrix}$

$\hat{Y}^T$-2 marks $\begin{bmatrix} 2.8 & 5.1 & 7.4 & 9.7 & 12 \\ -2 & 0 & 2 & 4 & 6 \end{bmatrix}$

$\hat{Y}^T \hat{Y}$- 2 marks $\begin{bmatrix} 326.7 & 120 \\ 120 & 60 \end{bmatrix}$

$\hat{\varepsilon}^T - 2\ marks$ $\begin{bmatrix} -0.8 & 0.9 & 0.6 & -0.7 & 0 \\ 0 & -2 & 2 & 2 & -2 \end{bmatrix}$

$\hat{\varepsilon}^T \hat{\varepsilon} - 2\ marks$ $\begin{bmatrix} 2.3 & -2 \\ -2 & 16 \end{bmatrix}$

$Y^T Y$- 2 marks $\begin{bmatrix} 329 & 118 \\ 118 & 76 \end{bmatrix}$

$Y^T Y = \hat{Y}^T \hat{Y} + \hat{\varepsilon}^T \hat{\varepsilon}$-2 marks $\begin{bmatrix} 329 & 118 \\ 118 & 76 \end{bmatrix} = \begin{bmatrix} 326.7 & 120 \\ 120 & 60 \end{bmatrix} + \begin{bmatrix} 2.3 & -2 \\ -2 & 16 \end{bmatrix}$

2. Considers the following eigenvalues (475, 6.09, 0.03, 0.82) and eigen vectors $\begin{bmatrix} 0.01 & 0.18 & -0.71 & 0.68 \\ 0.02 & 0.18 & 0.71 & 0.69 \\ 0.03 & 0.67 & 0 & -0.25 \\ -0.99 & 0.04 & 0.01 & 0.01 \end{bmatrix}$. Apply Spectral Decomposition theorem and prove your of selection of number of components. (10)

formula 1 marks

$$\sum_{i=1}^{d} \boldsymbol{\lambda_i \vec{e}_i \vec{e}^{i^T}} \cong \sum_{i=1}^{p} \lambda_i \vec{e}_i\, \vec{e}^{i^T}$$

PVE(v1)=98.56, PVE(v2)=1.26 --- 2 marks

LHS

$\lambda_1 \vec{e}_1 \vec{e}_1{}^T + \lambda_2 \vec{e}_2 \vec{e}_2{}^T + \lambda_3 \vec{e}_3 \vec{e}_3{}^T + \lambda_4 \vec{e}_4 \vec{e}_4{}^T$- 4 marks

RHS $\lambda_1 \vec{e}_1 \vec{e}_1{}^T + \lambda_2 \vec{e}_2 \vec{e}_2{}^T$ - 2marks

Since LHS and RHS are approximately equal selection of 2 components are justified – 1 mark

b) In the process of separation of population, apply the maximum ratio formula and find the college in which the student can be admitted with his given CGPA and GRE scores $X_0^T = [8.5, 318]$.

$$\overline{X_1} = \begin{bmatrix}7.2\\295\end{bmatrix};\ \overline{X_2} = \begin{bmatrix}8.2\\312\end{bmatrix};\ \overline{X_3} = \begin{bmatrix}8.8\\323\end{bmatrix}; S_{Pooled} = \begin{bmatrix}0.96 & 0.32\\0.32 & 3.87\end{bmatrix} \quad (10)$$

$$S_{pooled}^{-1} = \begin{bmatrix}1.07 & -0.09\\-0.09 & 0.27\end{bmatrix}$$ -4 marks

$$D_1^2 = [1.3 \quad 23]\begin{bmatrix}1.07 & -0.09\\-0.09 & 0.27\end{bmatrix}\begin{bmatrix}1.3\\23\end{bmatrix}=139.26$$ - 2 marks

$$D_2^2 = [1.3 \quad 23]\begin{bmatrix}1.07 & -0.09\\-0.09 & 0.27\end{bmatrix}\begin{bmatrix}1.3\\23\end{bmatrix}=139.26$$ - 2 marks

$$D_3^2 = [1.3 \quad 23]\begin{bmatrix}1.07 & -0.09\\-0.09 & 0.27\end{bmatrix}\begin{bmatrix}1.3\\23\end{bmatrix}=139.26$$ - 2 marks

3.Calculate Fisher discriminant score for the data $X_0^T = [-3, -2]$ and allocate it to the appropriate groups. n1=n2=n3=3. p1=p2=0.2,p3=0.6.

$$\text{Low:} \quad X_1 = \begin{bmatrix}-3 & 4\\-1 & 2\\-2 & 0\end{bmatrix};\ \overline{X_1} = \begin{bmatrix}-2\\2\end{bmatrix}; S_1 = \begin{bmatrix}1 & -1\\-1 & 4\end{bmatrix}$$

$$\text{Med:} \quad X_2 = \begin{bmatrix}-1 & 5\\1 & 3\\0 & 1\end{bmatrix};\ \overline{X_2} = \begin{bmatrix}0\\3\end{bmatrix}; S_2 = \begin{bmatrix}1 & -1\\-1 & 4\end{bmatrix}$$

$$\text{High:} \quad X_3 = \begin{bmatrix}1 & -2\\0 & 0\\-1 & -4\end{bmatrix};\ \overline{X_3} = \begin{bmatrix}0\\-2\end{bmatrix}; S_3 = \begin{bmatrix}1 & 1\\1 & 4\end{bmatrix} \quad (20)$$

$$S_{pooled}^{-1} = \frac{1}{35}\begin{bmatrix}36 & 3\\3 & 9\end{bmatrix}$$ - 4 marks

$$S_{pooled}^{-1} = \begin{bmatrix}1.03 & 0.09\\0.09 & 0.26\end{bmatrix}$$

Π 1: 5 mark

$$\bar{X}_1^T S_{pooled}^{-1} = [-2 \quad 2]\frac{1}{35}\begin{bmatrix}36 & 3\\3 & 9\end{bmatrix} = \frac{1}{35}[-66 \quad 12]$$

$$\bar{X}_1^T S_{pooled}^{-1} = [-1.88 \quad 0.34]$$

$$\bar{X}_1^T S_{pooled}^{-1}\bar{X}_1 = \frac{1}{35}[-66 \quad 12]\begin{bmatrix}-2\\2\end{bmatrix} = \frac{156}{35}$$

$$\bar{X}_1^T S_{pooled}^{-1}\bar{X}_1 = [3.77 \quad 0.69] = 4.46$$

$$\hat{d}_1(x_0) = \ln(0.2) + \frac{-66}{35}(-3) + \frac{12}{35}(-2) - \frac{1}{2}\left(\frac{156}{35}\right)$$
=-1.6+5.66-0.69-2.23=1.14

$$\boldsymbol{\hat{d}_1(x_0) = 1.14}$$

Π 2: 5 mark

$$\bar{X}_2^T S_{pooled}^{-1} = [0 \quad 3]\frac{1}{35}\begin{bmatrix}36 & 3\\3 & 9\end{bmatrix} = \frac{1}{35}[9 \quad 27]$$

$$\bar{X}_2^T S_{pooled}^{-1} = [0.26 \quad 0.77]$$

$$\bar{X}_2^T S_{pooled}^{-1}\bar{X}_1 = \frac{1}{35}[9 \quad 27]\begin{bmatrix}0\\3\end{bmatrix} = \frac{81}{35}$$

$$\bar{X}_2^T S_{pooled}^{-1}\bar{X}_2 = [0.26 \quad 0.77]\begin{bmatrix}0\\3\end{bmatrix} = 2.31$$

$$\hat{d}_2(x_0) = \ln(0.2) + \frac{9}{35}(-3) + \frac{27}{35}(-2) - \frac{1}{2}\left(\frac{81}{35}\right)$$
=-1.6-0.77-1.54-1.15=-5.06

$$\boldsymbol{\hat{d}_2(x_0) = -5.06}$$

Π 3: 5 mark

$\bar{X}_3^T S_{pooled}^{-1} = [0 \quad -2]\frac{1}{35}\begin{bmatrix}36 & 3\\ 3 & 9\end{bmatrix} = \frac{1}{35}[-6 \quad -18]$ $\bar{X}_2^T S_{pooled}^{-1} = [-0.18 \quad -0.52]$

$\bar{X}_3^T S_{pooled}^{-1}\bar{X}_1 = \frac{1}{35}[-6 \quad -18]\begin{bmatrix}0\\ -2\end{bmatrix} = \frac{36}{35}$ $\bar{X}_2^T S_{pooled}^{-1}\bar{X}_2 = [-0.18 \quad -0.52]\begin{bmatrix}0\\ -2\end{bmatrix} = 1.04$

$$\hat{d}_3(x_0) = \ln(0.6) + \frac{-6}{35}(-3) + \frac{-18}{35}(-2) - \frac{1}{2}\left(\frac{36}{35}\right)$$

=-0.51+0.51+1.03-0.51=0.52

$$\boldsymbol{\hat{d}_3(x_0) = 0.52}$$

Since $\hat{d}_1(x_0)$ greater assign data to π-2 --- 1 mark

PART B Answer all the questions 10 marks

4. Derive the steps of calculating Principal Components. (10)

1. Let $\bar{X}$ be the **mean** vector (taking the mean of all rows)

   X=$a_1, a_2, \ldots a_d$ set of N×d vectors and let $\bar{X}$ be their average of d attributes.

   $$X = \begin{bmatrix} a_{11} & \ldots & a_{1d} \\ . & \ldots & . \\ a_{N1} & \ldots & a_{Nd} \end{bmatrix} \qquad \bar{X} = \frac{1}{N}\sum_{i=1}^{i=N} a_{ij},\ j = 1\ to\ d$$

2. Adjust the original data by the mean X' = X – $\bar{X}$
3. Compute the **covariance** matrix C of adjusted X

   $$C = \frac{1}{N-1}XX^T = \frac{1}{N-1}\begin{bmatrix}(a_1 - \bar{X})^T \\ (a_2 - \bar{X})^T \\ \vdots \\ . \\ (a_d - \bar{X})^T\end{bmatrix}[a_1 - \bar{X} \quad a_2 - \bar{X} \quad \ldots\ a_d - \bar{X}]$$

   And C=$\begin{bmatrix} x_{11} & \ldots & x_{1d} \\ . & \ldots & . \\ x_{d1} & \ldots & x_{dd} \end{bmatrix}$

   C is square, symmetric, Covariance matrix - Normalized centered data matrix
4. Find the **eigenvectors and eigenvalues** of C.

   In computational terms the principal components are found by calculating the eigenvectors and eigenvalues of the data covariance matrix. This process is equivalent to finding the axis system in which the co-variance matrix is diagonal. The eigenvector with the largest eigenvalue is the direction of greatest variation, the one with the second largest eigenvalue is the (orthogonal) direction with the next highest variation and so on. To see how the computation is done we will give a brief review on eigenvectors/eigenvalues. The eigen values of C are defined as the roots of :

   determinant (C-λI)=|(C- λI)|=0

   $$\left\|\begin{bmatrix} x_{11} & \ldots & x_{1d} \\ . & \ldots & . \\ x_{d1} & \ldots & x_{dd} \end{bmatrix} - \lambda\begin{bmatrix} 1 & \ldots & 0 \\ . & \ldots & . \\ 0 & \ldots & 1 \end{bmatrix}\right\| = 0$$

   $$\left\|\begin{bmatrix} x_{11} & \ldots & x_{1d} \\ . & \ldots & . \\ x_{d1} & \ldots & x_{dd} \end{bmatrix} - \begin{bmatrix} \lambda_{11} & \ldots & 0 \\ . & \ldots & . \\ 0 & \ldots & \lambda_{dd} \end{bmatrix}\right\| = 0$$

   – I is a d×d identity matrix, yields a polynomial (characteristic polynomial) (degree d and has d roots)

- λ be the eigenvalue (roots) of C. Then here exists a vector $\vec{e}$ such that

$$C\,\vec{e} = \lambda\,\vec{e}$$

$$(C-\lambda I)\,\vec{e}=0$$

$$\left[\begin{bmatrix} x_{11} & \cdots & x_{1d} \\ . & \cdots & . \\ x_{d1} & \cdots & x_{dd} \end{bmatrix}\right] - \begin{bmatrix} \lambda_{11} & \cdots & 0 \\ . & \cdots & . \\ 0 & \cdots & \lambda_{dd} \end{bmatrix}\begin{bmatrix} e_1 \\ .. \\ e_d \end{bmatrix} = 0$$

The vector $\vec{e}$ be an eigenvector of C associated with eigenvalue λ. Notice that there is no unique solution for $\vec{e}$ in the above equation. It is a direction vector only and can be scaled to any magnitude. To find the numerical solution of $\vec{e}$ we need to set one of its elements to an arbitrary value, say 1, which gives us a set of simultaneous equations to solve for the other elements.
- Solve $(C-\lambda I)\,\vec{e}=0$ for each λ (λ_1, λ_1 .. λ_d, -d roots) to obtain eigenvectors $\vec{e}_{\,i}$
- $\vec{e}_{\,i} = \vec{e}_{\,i1}, \vec{e}_{\,i2} .. \vec{e}_{\,id}$ are d ×d orthonormal vectors
- *eigenvectors* of *C* is $\vec{e}$ such that $C\vec{e}=\lambda\vec{e}$,
- $C\vec{v}=\lambda\vec{e} \Leftrightarrow (C-\lambda I)\,\vec{e}=0$
- Expressing C interms of $\vec{e}_{\,i1}, \vec{e}_{\,i2} .. \vec{e}_{\,id}$ has not changed the size of the data. Eigen values λ_i corresponds to variance on each component i. Sort the eigenvectors $\vec{e}_{\,i}$ according to their eigenvalue:

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$$

The new data set can be calculation:

$$\begin{bmatrix} y_{11} & \cdots & y_{1d} \\ . & \cdots & . \\ y_{N1} & \cdots & y_{Nd} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1d} \\ . & \cdots & . \\ a_{N1} & \cdots & a_{Nd} \end{bmatrix}\begin{bmatrix} \vec{e}_{\,i1} \\ \vec{e}_{\,i2} \\ \cdots \\ \vec{e}_{\,id} \end{bmatrix}, \qquad i = 1\ to\ d$$

Consider a linear combinations in which the new data is calculated. Take the first *p* components based on top *p* eigenvectors $\vec{e}_{\,i}$. These are the directions with the largest variances.

$$Y_1 = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1d}X_d$$
$$Y_2 = e_{21}X_1 + e_{22}X_2 + \cdots + e_{2d}X_d$$
$$.$$
$$.$$
$$Y_p = e_{d1}X_1 + e_{d2}X_2 + \cdots + e_{dd}X_d$$

Each of these can be thought of as a linear regression, predicting Y_i from $X_1, X_2, \ldots, X_d$. There is no intercept, but $e_{i1}, e_{i2}, \ldots, e_{id}$ can be viewed as regression coefficients.

The Principal Components selection:

$$PVE\_PC_i = (\lambda_i / \textstyle\sum_{i=1}^{d} \lambda_i) \times 100$$

Where PVE-Percentage of Variance Explained

$$PVE\_PC_1 \geq PVE\_PC_2 \geq \cdots PVE\_PC_p \geq \cdots \geq PVE\_PC_d$$