Q1. (12 pts in total)

Step-1: For each dimension i, calculate the mean e_i (expected value); and for each L-dimensional data $\{x_1, ..., x_L\}$ find $\{x_1 - e_1, ..., x_L - e_L\}$

mean vector $(e_1, e_2)^T = (\frac{2+4+1+5}{4}, \frac{2+4+5+1}{4})^T = (3,3)^T$

For data (2,2), difference from mean vector $= (2-3, 2-3)^T = (-1,-1)^T$

For data (4,4), difference from mean vector $= (4-3, 4-3)^T = (1,1)^T$

For data (1,5), difference from mean vector $= (1-3, 5-3)^T = (-2,2)^T$

For data (5,1), difference from mean vector $= (5-3, 1-3)^T = (2,-2)^T$

Step-2: Obtain the covariance matrix Σ.

$$Y = \begin{pmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{pmatrix}$$

$$\Sigma = \frac{1}{4}YY^T = \frac{1}{4}\begin{pmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ 1 & 1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix} = \begin{pmatrix} \frac{5}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

Step-3: Find the eigenvalues and eigenvectors of Σ, choose the eigenvectors of unit lengths

$$\begin{vmatrix} \frac{5}{2}-\lambda & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-\lambda \end{vmatrix} = 0 \to (\frac{5}{2}-\lambda)^2 - (-\frac{3}{2})^2 = 0 \to \lambda = 4 \ or \ \lambda = 1$$

When $\lambda = 4$,

$$\begin{pmatrix} \frac{5}{2}-4 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to \begin{pmatrix} -\frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & -\frac{3}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to x_1 + x_2 = 0$$

We choose the eigenvector of unit length $(x_1, x_2)^T = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$

When $\lambda = 1$,

$$\begin{pmatrix} \frac{5}{2}-1 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to \begin{pmatrix} \frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{3}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to x_1 - x_2 = 0$$

We choose the eigenvector of unit length $(x_1, x_2)^T = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T$

Step-4: Arrange the eigenvectors in descending order of eigenvalues

$$\Phi = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Step-5: Transform the given L-dimensional vectors by eigenvector matrix $Y = \Phi^T X$

For data (2,2),

$$Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2.83 \end{pmatrix}$$

For data (4,4),

$$Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 5.66 \end{pmatrix}$$

For data (1,5),

$$Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 2.83 \\ 4.24 \end{pmatrix}$$

For data (5,1),

$$Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} -2.83 \\ 4.24 \end{pmatrix}$$

Step-6: For each "transformed" L-dimensional vector, keep only the K values $\{y_1, ..., y_k\}$ corresponding to the smallest K eigenvalues

(2,2)$\Longrightarrow$(2.83)

(4,4)$\Longrightarrow$(5.66)

(1,5)$\Longrightarrow$(4.24)

(5,1)$\Longrightarrow$(4.24)

Suggested Marking Scheme: 2 pts for correctly showing each step.

Q2. (8 pts in total)

(a) c (2 pts) and d (2 pts) are outliers for the given setting.

Reasoning:

- a's ϵ-neighborhood is $\{a, b, d, e, f\}$
- b's ϵ-neighborhood is $\{a, b, c, e, f\}$
- c's ϵ-neighborhood is $\{b, c, f\}$
- d's ϵ-neighborhood is $\{a, d\}$
- e's ϵ-neighborhood is $\{a, b, e, f\}$
- f's ϵ-neighborhood is $\{a, b, c, e, f\}$

Since the size threshold is 3, then c and d should be outliers.

Note: The question does not ask for an explanation. 2 points could be granted if there is an partially-correct explanation.

(b)

$$
\begin{aligned}
LOF(d) &= \frac{\sum_{o \in N_3(d)} \frac{lrd_3(o)}{lrd_3(d)}}{3} \\
&= \frac{1}{3 \times lrd_3(d)}(lrd_3(a) + lrd_3(c) + lrd_3(b)) \\
&= \frac{1}{3 \times \frac{1}{6}}(\frac{1}{4} + \frac{1}{5} + \frac{1}{3}) = \frac{47}{30} \approx 1.57(\text{2 pts})
\end{aligned}
$$

$$
\begin{aligned}
LOF(c) &= \frac{\sum_{o \in N_3(c)} \frac{lrd_3(o)}{lrd_3(c)}}{3} \\
&= \frac{1}{3 \times lrd_3(c)}(lrd_3(a) + lrd_3(b) + lrd_3(f)) \\
&= \frac{1}{3 \times \frac{1}{5}}(\frac{1}{4} + \frac{1}{3} + \frac{1}{3}) = \frac{47}{30} \approx 1.53(\text{2 pts})
\end{aligned}
$$

Note: If the final number is not correct, 3 points should be granted if the equation used is correct