CMSC5724: Quiz 3

Name: **Student ID:**

Problem 1 (50%). Run the Apriori algorithm on the following set of transactions to find itemsets with frequencies at least 3. You only need to give (i) the candidate itemsets at the beginning of each round, and (ii) the frequent itemsets found at each round.

transaction ID	items
1	A, B, D, E, F
2	A, B, D, F
3	A, B, F
4	C, D, F
5	B, C, E
6	A, C

Answer: The Apriori algorithm generates:

$$\begin{array}{ll}
C_1: & \{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\} \\
F_1 & \{A\}, \{B\}, \{C\}, \{D\}, \{F\} \\
C_2: & \{A,B\}, \{A,C\}, \{A,D\}, \{A,F\}, \{B,C\}, \{B,D\}, \{B,F\}, \{C,D\}, \{C,F\}, \{D,F\} \\
F_2 & \{A,B\}, \{A,F\}, \{B,F\}, \{D,F\} \\
C_2: & \{A,B,F\} \\
F_2 & \{A,B,F\}
\end{array}$$

Problem 2 (50%) Consider a set S of three 2D points: $\{(1,4),(1,1),(4,1)\}$. Use PCA to find the line passing the origin on which the projections of the points in S have the largest variance.

Answer: After aligning the geometry center of S to the origin, we obtain the following set of points: $\{(-1,2),(-1,-1),(2,-1)\}$. Its covariance matrix is

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Solve the characteristic equation of A:

$$\begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} = 0 \Rightarrow$$
$$(2-\lambda)^2 - 1 = 0$$

Solving the equation gives the two eigenvalues of A: $\lambda_1 = 3$ and $\lambda_2 = 1$. The line sought in the question is determined by an eigenvector of λ_1. In other words, we want to find $\begin{bmatrix} x \\ y \end{bmatrix}$ satisfying:

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 3 \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{1}$$

Setting $x = 1, y = -1$ satisfies the equation. Therefore, the line is the one passing the origin and the point $(1,-1)$.