

Data Mining & Analytics
Assignment-1

Sanjay S

126018042

① Feature	E ₁	E ₂	E ₃	E ₄	E ₅
x	2.3	0.85	2.2	1.9	3.1
y	2.4	0.7	2.9	2.2	3.0

No. of feature $\Rightarrow n = 2$ No. of Sample $N = 4$

$$\text{Mean } \bar{x} = \frac{2.3 + 0.85 + 2.2 + 1.9}{4} = 2.04$$

$$\bar{y} = 2.24$$

Cov. matrix =

$$\text{cov}(x, x) = \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2 = 0.938$$

$$\text{cov}(x, y) = 0.8408$$

$$\text{cov}(y, y) = 0.833$$

$$C = \begin{bmatrix} 0.938 & 0.8408 \\ 0.8408 & 0.833 \end{bmatrix}$$

$$\det(C - \lambda I) = 0$$

$$\lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$S - \lambda I = 0$$

$$\begin{bmatrix} 0.938 - \lambda & 0.8408 \\ 0.8408 & 0.833 - \lambda \end{bmatrix} = 0$$

$$\lambda^2 - 1.791\lambda + 0.094 = 0$$

$$\lambda = 1.791 \pm \sqrt{(1.791)^2 - 4(0.094)}$$

$$\lambda_1 = 1.736, \quad \lambda_2 = 0.054$$

$$(S - \lambda_1 I) v = 0$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\frac{U_1}{0.843} = \frac{U_2}{0.788} = 1$$

$$U = \begin{bmatrix} 0.8405 \\ 0.7915 \end{bmatrix}$$

variance, PC1 has variance of 96.99%,
PC2 has variance of 8.01%.

Projection onto PC1

$$E_1 : 0.4436 E_2 : -0.1772 E_3 : 0.5707$$

$$E_4 : -0.1290 E_5 : 1.2919$$

final PCA result $\begin{bmatrix} 0.7847 \\ 0.6890 \end{bmatrix}$ has 1D projections or
 $\begin{bmatrix} 0.4436 & -0.1772 & 0.5707 & -0.1290 & 1.2919 \end{bmatrix}$ and
 variance of 97%.

v) A db has 8 transactions with min-sup = 25%.
max-cof = 60%.

tr id

Items

T₁ {A, B, C, D}

T₂ {A, C, D, F, G}

T₃ {C, D, F, G, A}

T₄ {A, D, F, B}

T₅ {B, G, C}

T₆ {D, F, G}

T₇ {A, B, G}

T₈ {C, D, F, G}

No. of transaction

$$\Rightarrow 8 \times 0.25$$

$$= 2 (2)$$

frequent 1 items	count	frequent 2 items	count
A	5	(A, B)	3
B	4	(A, C)	3
C	5	(A, D)	4
D	6	(A, G)	2
E	1	(B, C)	2
F	4	(B, D)	2
G	5	(B, G)	2

frequent 3 items	count		
(A, B, D)	2	(C, F)	2
(A, C, D)	3	(C, G)	3
(A, D, F)	2	(D, F)	4
(C, C, D, F)	2	(D, G)	3
(C, C, D, G)	2	(F, G)	2
(C, D, F, G)	2		

(i) Itemset $\{A, B, D\}$ ($\text{sup}=2$)

$$(A, B, D) \rightarrow A, D \Rightarrow 2/3 \Rightarrow 66.67\% \text{ (strong)}$$

$$(A, D) \rightarrow B \Rightarrow 2/4 \Rightarrow 50\% X$$

$$(B, D) \rightarrow A \Rightarrow 2/2 \Rightarrow 100\% \nearrow$$

$$A \rightarrow (B, D) \Rightarrow 2/5 \Rightarrow 40\% X$$

$$B \rightarrow (A, D) \Rightarrow 2/4 \Rightarrow 50\% X$$

$$D \rightarrow (A, B) \Rightarrow 2/6 \Rightarrow 33.33\% X$$

$\{A, C, D\}$ ($\text{sup}=3$)

$$A \rightarrow \{A, C, D\} = 60\% \nearrow$$

$$C \rightarrow \{A, C, D\} = 60\% \nearrow$$

$$D \rightarrow \{A, C, D\} = 50\% X$$

$$(A, C) \rightarrow D \Rightarrow 100\% \nearrow$$

$$(A, D) \Rightarrow C \Rightarrow 75\% \nearrow$$

$$(C, C, D) \Rightarrow A \Rightarrow 75\% \nearrow$$

$\{A, D, F\}$ ($\text{sup}=2$)

$$A \rightarrow \{D, F\} \Rightarrow 2/15 \Rightarrow 40\% X \quad (A, D) \rightarrow F = 50\% X$$

$$D \rightarrow \{A, F\} \Rightarrow 33.33\% X$$

$$(A, F) \rightarrow D = 100\% \nearrow$$

$$F \rightarrow \{A, D\} \Rightarrow 50\% X$$

$$(D, F) \rightarrow A = 50\%$$

$C_1 D_1 F \supset 2$

$C \rightarrow C_1 D_1 F \supset 40\%$ X

$D \rightarrow C_1 F \supset 33.33\%$ X

$F \rightarrow A_1 D \supset 50\%$ X

$(C_1 D) \rightarrow F \Rightarrow 50\%$ X

$(C_1 F) \rightarrow D \Rightarrow 100\%$ X

$(D_1 F) \rightarrow C \Rightarrow 50\%$ X

$C_1 D_1 G \supset 2$

$C \rightarrow C_1 D_1 G \supset 40\%$ X

$D \rightarrow C_1 G \supset 33.33\%$ X

$G \rightarrow C_1 D \supset 40\%$ X

$(C_1 D) \rightarrow G \supset 50\%$ X

$(C_1 G) \rightarrow G \supset 66.67\%$ X

$(D_1 G) \rightarrow G \supset 66.67\%$ X

Strong Annotation Rules

$(A_1 B) \rightarrow D (66.67\%)$

$(B_1 D) \rightarrow A (100\%)$

$A \rightarrow C_1 D \supset (60\%)$

$C \rightarrow A_1 D \supset (60\%)$

$(A_1 C) \rightarrow D (100\%)$

$(A_1 D) \rightarrow C (75\%)$

$(C_1 D) \rightarrow A (75\%)$

$(A_1 F) \rightarrow D (100\%)$

$(C_1 F) \rightarrow D (100\%)$

$(C_1 G) \rightarrow D (66.67\%)$

$(D_1 G) \rightarrow C (66.67\%)$

$(D_1 G) \rightarrow F (66.67\%)$

$(F_1 G) \rightarrow D (100\%)$

2) FP Growth Algorithm
 min: threshold for support
 \Rightarrow 50% , confidence = 60%,
 min threshold $\geq 50/100 \times 6$
 $\Rightarrow 3$

Transaction	List of Items
T ₁	l ₁ , l ₂ , l ₃
T ₂	l ₂ , l ₃ , l ₄
T ₃	l ₄ , l ₅
T ₄	l ₁ , l ₂ , l ₄
T ₅	l ₁ , l ₂ , l ₃ , l ₅
T ₆	l ₁ , l ₂ , l ₃ , l ₄

frequent 1 itemset count

l ₁	4
l ₂	5
l ₃	4
l ₄	4
l ₅	2

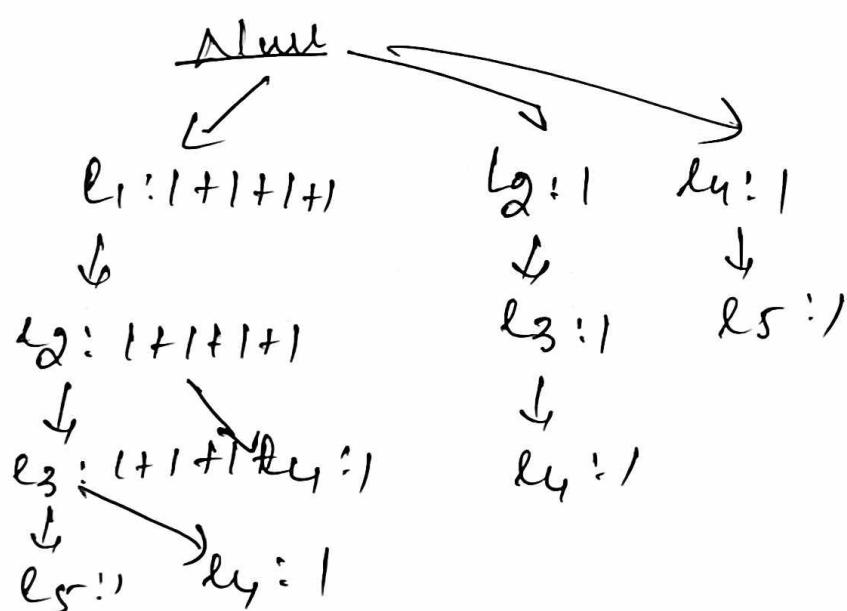
frequent 2 itemset count

(l ₁ , l ₂)	4
(l ₁ , l ₃)	3
(l ₂ , l ₃)	4
(l ₂ , l ₄)	3

frequent 3 itemset count

(l ₁ , l ₂ , l ₃)	3
---	---

No more frequent pattern.



Items conditional pattern base

l_5 $\{ \{ l_1, l_2, l_3 : 1 \} \cup \{ l_4 : 1 \} \}$

l_4 $\{ \{ l_1, l_2, l_3 : 1 \}, \{ l_1, l_2 : 1 \}, \{ l_2, l_3 : 1 \} \}$

l_3 $\{ \{ l_1, l_2 : 3 \}, \{ l_2 : 1 \} \}$

l_2 $\{ \{ l_1, l_4 \} \}$

l_1 —

<u>Item</u>	<u>condition-frequent pattern</u>	<u>Frequent pattern generated</u>
l_4	$\{ l_2 : 3 \}$	$\{ \{ l_2, l_4 \} \}$
l_3	$\{ l_2 : 4 \}$	$\{ \{ l_2, l_3 \} \}$
l_2	$\{ l_4 : 4 \}$	$\{ \{ l_2, l_4 \} \}$
l_1	—	

Frequent