

# 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  $N=5$

$$\text{Mean } \bar{x} = \cancel{2.5} 2.04$$

$$\bar{y} = 2.24$$

Cov. matrix  $\Rightarrow$

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

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

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

$$C = \begin{bmatrix} 0.938 & 0.8403 \\ 0.8405 & 0.833 \end{bmatrix}$$

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

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

$$S - \lambda I = 0$$

$$\begin{bmatrix} 0.938 - \lambda & \cancel{0.840} 0.840 \\ 0.8401 & 0.833 - \lambda \end{bmatrix} = 0$$

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

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

$$\lambda_1 = 1.736, \lambda_2 = 0.0548$$

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

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

$$\frac{u_1}{0.843} = \frac{u_2}{0.788} = 1$$

$$u = \begin{pmatrix} 0.8405 \\ 0.7915 \end{pmatrix}$$

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

Projection onto PC1

$$E_1: 0.4436 \quad E_2: -2.1772 \quad E_3: 0.5707$$

$$E_4: -0.1290 \quad E_5: 1.2919$$

Final PCA result  $\begin{bmatrix} 0.7247 \\ 0.6890 \end{bmatrix}$  has 1D projection or

$[0.4436 \quad -2.1772 \quad 0.5707 \quad -0.1290 \quad 1.2919]$  and  
variance of 97%.

✓) A db has 8 transactions with min-sup = 25%,  
max-conf = 60%.

tr id

Items

T<sub>1</sub>

{A, B, C, D}

T<sub>2</sub>

{A, C, D, F}

T<sub>3</sub>

{C, D, F, G, A}

T<sub>4</sub>

{A, D, F, B}

T<sub>5</sub>

{B, G, C}

T<sub>6</sub>

{D, F, G}

T<sub>7</sub>

{A, B, G}

T<sub>8</sub>

{C, D, F, G}

No. of transactions

$$\Rightarrow 8 \times 0.25$$

$$= 2 (\geq 2)$$

(f) 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	(C, D)		2
(A, B, D)		2	(C, F)		4
(A, C, D)		3	(C, G)		2
(A, D, F)		2	(D, F)		3
(C, C, D, F)		2	(D, G)		3
(C, C, D, G)		2	(F, G)		2
(C, D, F, G)		2			

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

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

$\{A, D\} \rightarrow B \Rightarrow 2/4 \Rightarrow 50\%$  X

$\{B, D\} \rightarrow A \Rightarrow 2/2 \Rightarrow 100\%$  ✓

$A \rightarrow \{B, D\} \Rightarrow 2/5 \Rightarrow 40\%$  X

$B \rightarrow \{A, D\} \Rightarrow 2/4 \Rightarrow 50\%$  X

$D \rightarrow \{A, B\} \Rightarrow 2/6 \Rightarrow 33.34\%$  X

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

$A \rightarrow \{C, D\} = 60\%$  ✓

$C \rightarrow \{A, D\} = 60\%$  ✓

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

$\{A, C\} \rightarrow D \Rightarrow 100\%$  ✓

$\{A, D\} \rightarrow C \Rightarrow 75\%$  ✓

$\{C, D\} \rightarrow A \Rightarrow 75\%$  ✓

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

$A \rightarrow \{D, F\} \Rightarrow 2/5 \Rightarrow 40\%$  X

$D \rightarrow \{A, F\} \Rightarrow 2/6 \Rightarrow 33.34\%$  X

$F \rightarrow \{A, D\} \Rightarrow 2/4 \Rightarrow 50\%$  X

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

$\{A, F\} \rightarrow D = 100\%$  ✓

$\{D, F\} \rightarrow A = 50\%$

$\angle C|D, F \text{ sup } 2$

$C \rightarrow \angle D, F \Rightarrow 40\% \quad X$

$D \rightarrow \angle C, F \Rightarrow 33.33\% \quad X$

$F \rightarrow \angle A, D \Rightarrow 50\% \quad X$

$(C|D) \rightarrow F \Rightarrow 50\% \quad X$

$(C, F) \rightarrow D \Rightarrow 100\% \quad \checkmark$

$(D, F) \rightarrow C \Rightarrow 50\% \quad X$

$\angle C|D, G \text{ sup } = 2$

$C \rightarrow \angle D, G \Rightarrow 40\% \quad X$

$D \rightarrow \angle C, G \Rightarrow 33.33\% \quad X$

$G \rightarrow \angle C, D \Rightarrow 40\% \quad X$

$(C|D) \rightarrow G \Rightarrow 50\% \quad X$

$(C, G) \rightarrow D \Rightarrow 66.67\% \quad \checkmark$

$(D, G) \rightarrow C \Rightarrow 66.67\% \quad \checkmark$

### Strong Association Rules

$(A|B) \rightarrow D (66.67\%)$

$(B|D) \rightarrow A (100\%)$

$A \rightarrow \angle C, D (60\%)$

$C \rightarrow \angle A, D (60\%)$

$(A|C) \rightarrow D (100\%)$

$(A|D) \rightarrow C (75\%)$

$(C|D) \rightarrow A (75\%)$

$(A|F) \rightarrow D (100\%)$

$(C|F) \rightarrow D (100\%)$

$(C, G) \rightarrow D (66.67\%)$

$(D, G) \rightarrow C (66.67\%)$

$(D, G) \rightarrow F (66.67\%)$

$(F, G) \rightarrow D (100\%)$

2) FP Growth Algorithm  
 min: threshold for support  
 $\Rightarrow 50\%$ , confidence = 60%

min threshold  $\Rightarrow 50/100 \times 6$   
 $\Rightarrow 3$

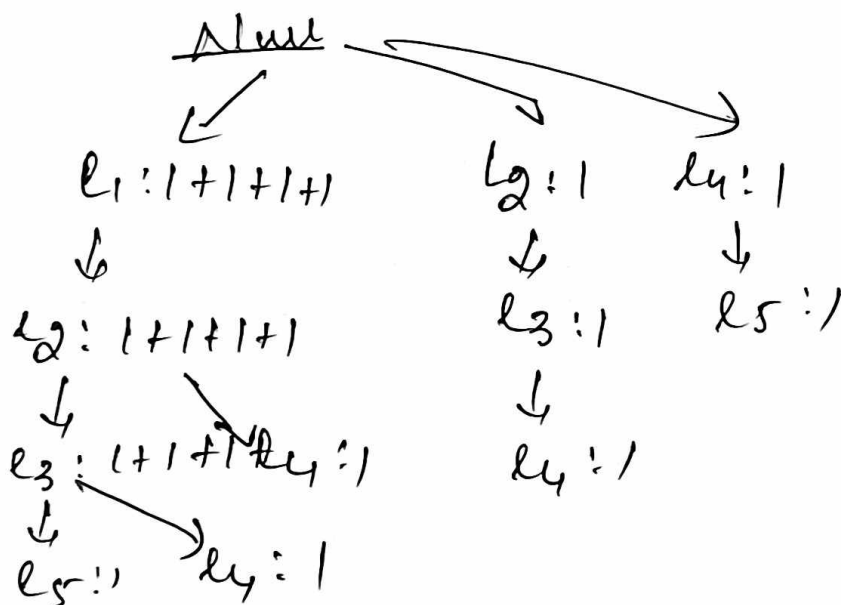
Transaction	List of items
T1	$l_1, l_2, l_3$
T2	$l_2, l_3, l_4$
T3	$l_4, l_5$
T4	$l_1, l_2, l_4$
T5	$l_1, l_2, l_3, l_5$
T6	$l_1, l_2, l_3, l_4$

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 patterns.





Items

conditional pattern base

$l_5$

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

$l_4$

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

$l_3$

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

$l_2$

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

$l_1$

Item

conditional frequent pattern

Frequent pattern generated

$l_4$

$\{ l_2 : 3 \}$

$\{ l_2, l_4 : 3 \}$

$l_3$

$\{ l_2 : 4 \}$

$\{ l_2, l_3 : 4 \}$

$l_2$

$\{ l_4 : 4 \}$

$\{ l_2, l_4 : 4 \}$

$l_1$

—

Frequent