k-Nearest Neighbour :-

	X_1	X_2	Y
1.	7	7	Good
2.	7	4	Bad
3.	3	4	Good
4.	1	4	Bad
	3	7	?

Assume $k = 3$

(7,7) (3,7)

$(7-3)^2 = 4^2 = 16$.

(7,4) (3,7)

$(7-3)^2 + (4-7)^2 = 4^2 + 3^2 = 16 + 9 = 25$

(3,4) (3,7)

$(3-3)^2 + (4-7)^2 = 0^2 + 3^2 = 9$.

(1,4) (3,7)

$(1-3)^2 + (4-7)^2 = 2^2 + 3^2 = 4 + 9 = 13$.

3 - 1 Good

Speed	Agility	Draft
2.5	6.00	no
3.75	8.00	no
2.25	5.50	no
3.25	8.25	no
2.75	7.5	no
4.5	5.0	no
3.5	5.25	no
3.0	3.25	no
4.0	4.0	no
4.25	3.75	no
2.0	2.0	no
5.0	2.5	no
8.25	8.5	no
5.75	8.75	yes
4.75	6.25	yes
5.5	6.75	yes
5.25	9.5	yes
7.0	4.25	yes
7.5	8.0	yes
7.25	5.75	yes
6.75	3	?

k = 7
distance (6.75, 3)

Baye's Theorem :-

$$P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$$

h - hypothesis

D - trains data

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Week	Yes
D4	Rain	Mild	High	Week	Yes
D5	Rain	Cool	Normal	Week	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No
T1	Sunny	Cool	High	Strong	?

$P(No|Sunny) = \frac{3}{14}$ $\quad P(Yes|Sunny) = \frac{2}{14}$

$P(No|Cool) = \frac{1}{14}$ $\quad P(Yes|Cool) = \frac{3}{14}$

$P(No|High) = \frac{4}{14}$ $\quad P(Yes|High) = \frac{3}{14}$

$P(No|Strong) = \frac{3}{14}$ $\quad P(Yes|Strong) = \frac{3}{14}$

$P(\text{Yes}/\text{Sunny, Cool, High, Strong}) = P(\text{Sunny}/\text{Yes}) \times P(\text{Cool}/\text{Yes}) \times P(\text{High}/\text{Yes}) \times P(\text{Strg}/\text{Yes}) \times P(\text{Yes})$

$P(B/A) = \frac{P(A/B) \times P(B)}{P(A)}$

$= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$

$= 0.005$

$P(\text{No}/\text{Sunny, Cool, High, Strg})$

$= P(\text{Sunny}/\text{No}) \times P(\text{Cool}/\text{No}) \times P(\text{High}/\text{No}) \times P(\text{Strg}/\text{No}) \times$

$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$

$= 0.0205$

$P(\text{No}) > P(\text{Yes})$

The result is the instance belong to No Class.

age	income	student	creditrating	buyscom
<=30	high	no	fair	no
<=30	high	no	excellent	no
31..40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	excellent	yes
>40	low	yes	excellent	no
31..40	low	yes	fair	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	excellent	yes
<=30	medium	yes	excellent	yes
31..40	medium	no	fair	yes
31..40	high	yes	excellent	no
>40	medium	no	excellent	no

$P(no/age<=30, medium, yes, fair) = P(age<=30/no) \times P(medium/no) \times P(yes/no) \times P(fair/no) \times P(no)$

$= \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{14}$

$= 0.006$

$P(yes/age<=30, medium, yes, fair) = P(age<=30/yes) \times P(medium/yes) \times P(yes/yes) \times P(fair/yes) \times P(yes)$

$= \frac{2}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{5}{9} \times \frac{9}{14}$

$= 0.02$

Student id	Algorithms	Networks	C++	DBMS	Placed in
S1	A	A	C	B	Amazon
S2	B	B	C	C	Cisco
S3	B	C	C	A	TCS
S4	B	A	B	B	Cisco
S5	A	B	B	B	Amazon
S6	C	C	B	C	TCS
S7	C	B	A	B	Amazon
S8	B	C	A	C	TCS
S9	B	B	A	A	TCS
S10	C	C	B	C	TCS
S11	C	C	B	C	TCS
S12	B	A	C	B	Cisco
Sx	A	B	C	C	?

Laplacian

$P(\text{Amazon} \mid A,B,C,c) = P(A/\text{Amazon}) \times P(B/\text{Amazon}) \times P(C/\text{Amazon}) \times P(c/\text{Amazon})$

$= \frac{2}{4} \times \frac{3}{4} \times \frac{2}{4} \times \frac{1}{4} \times \frac{3}{12}$

$= 0$

$P(\text{Gsco} \mid A,B,C,c) = P(A \mid \text{Gsco})$

$= \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} \times \frac{2}{4} \times \frac{3}{12}$

$P(\text{Tcs} \mid A,B,C,c) = \frac{0}{7} \times \frac{1}{7} \times \frac{1}{7} \times \frac{1}{7} \times \frac{6}{12}$

$y = \theta_0 + \theta_1 x$

$\theta_1 = \frac{\sum (x-\bar{x})^{*}(y-\bar{y})}{\sum (x-\bar{x})^2}$

$\theta_0 = \bar{y} - \theta_1 \bar{x}$

X	Y	$x-\bar{x}$	$y-\bar{y}$	
23	272	-18.2	-117	2129.4
86	677	44.8	288	12902.4
14	210	-27.2	-179	4868.8
55	482	13.8	93	1283.4
28	304	-13.2	-85	1122
$\bar{X} = 41.2$	$\bar{Y} = 389$		3442.8	22306

$$\theta_1 = \frac{22306}{3442.8}$$

$$= 6.47$$

$$\theta_0 = 389 - 6.47 \times 41.2$$

$$= 122.43$$

$$Y = 122.43 + 6.47X$$

$X = 25 \quad Y = ?$

$$Y = 284.18$$

X	Y	$x-\bar{x}$. $y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
29	105			
7	18			
HH	167			
23	83			
35	129			

$\bar{x} = 27.6 \quad \bar{y} = 100.4$

$\theta_0 = 10.08$

$\theta_1 = 4$

$Y = -10.08 + 4X$

$\theta_0 = \bar{y} - \theta_1 \bar{x}$

$= 100.4 - 27.6 \times 4$

$= -10$

$\theta_1 = 0$

2.33

0.5

hlw $0.5 \times$

0.5 1 1.5

1 2 3

$\frac{1}{2}(1)^2 + (1.5)^2 = 0.58$

-0.2 to 0.2

Gradient

ID	Size	Floor	Broadband rate	(y) Rental price	Error $(y-y')^2$	Predicted
1	500	11	8	320	[illegible]	93.13
2	550	7	50	830	[illegible]	107.246
3	620	9	7	400	[illegible]	114.991
4	630	5	24	390	72419.32	19.41
5	665	8	100	385	82786.82	134.424
6	700	4	8	410	78393.84	130.011
7	770	10	7	480	118778.03	142.417
8	880	12	50	600	186565.52	168.076
9	920	14	8	570	159688.15	170.31
10	1000	9	24	620	181200.63	187.316

$\alpha = 0.0000002$ $w[0] = -0.146$ $w[1] = 0.185$

$w[2] = -0.044$ $w[2] = 0.119$

	error $y-y'$	error $*x_1$ for x_1	error $*x_2$ for x_2	error $*x_3$ for x_3
	$w[0]$			
1	226.74	113370	906.96	1813.92
2	272.59	149924.5	1908.13	13629.5
3	284.85	176607	2563.65	1993.95
4	270.79	170597.7	1353.95	6498.96
5	250.36	166489.4	2002.88	25036.42
6	279.69	195782.78	1118.76	2237.52
7	337.11	259570.96	3371.1	2359.74
8	421.68	379879.24	5180.17	21584.05
9	399.37	367423.83	5591.23	3194.99
10	432.42	432423.35	8891.81	10378.16
	3185.61	2412673.9	27818.65	88727.43

$w[j] = w[j] + \alpha \Sigma (D w_{ij})$

$w[0] = -0.146 + 0.0000002\,(3185.61)$

$= -0.145$

$w[1] = 0.185 + 0.0000002\,(24126\ 73.9)$

$= 0.2332$

$w[2] = -0.0434$

$w[3] = 0.1207$

A multivariate linear regression model has been built to predict the price in $ of a house based on the features size, no. of rooms, age describing the characteristic of the house. Assuming learning rate of 0.001 and initial weights $w[0]=400$, $w[1]=0.2$, $w[2]=0.5$, $w[3]=0.1$ Calculate weights at next iteration. using gradient descent algorithm

Size	No. of rooms	Size	Price	Predicted	Error
2105	5	45	840	828	12
1410	3	40	690	687.5	2.5
1550	3	30	720	714.5	5.5
900	2	26	600	583.6	16.4

for w_0	for w_1	for w_2	for w_3
12	25260	60	540
2.5	3525	1.5	100
5.5	8525	16.5	16.5
16.4	14760	32.8	426.4
	52070	116.8	1231.4

$w_0 = 400 + 0.001(36.4)$

$= 400.036$

$w_1 = 0.2 + 0.001(52070)$

$= 52.27$

$w_2 = 0.5 + 0.001(116.78)$

$= 0.6167$

$w_3 = 0.1 + 0.001(1208.4)$

$= 1.3314$

Consider a regression in which dependent variable is Hours

$w_0 = -60.0 \quad w_1 = 0.5 \quad w_2 = -0.1 \quad w_3 = 2.4 \quad \alpha = 0.001$

Income	Job	Family	Hours	Predicted	Error
120	40	5	8	8	0
150	30	3	15	19.2	-4.2

w_0	w_1	w_2	w_3
0	0	0	0
-4.2	-630	-126	-12.6
-4.2	-630	-126	-12.6

$w_0 \quad -60 + -4.2 \times 0.0001$

$= -60.00042$

$w_1 = 0.5 - 630 \times 0.0001$

$= -0.437$

$w_2 = -0.1 - 126 \times 0.000)$

$= -0.1126$

$w_3 = 2.4 - 12.6 \times 0.0001$

$= 2.39874$

A contigency table for binary data

		object j		
		1	0	sum
object i	1	q	r	q+r
	0	s	t	s+t
	sum	q+s	r+t	P

q represents no. of attributes which are equal to 1 for both objects i & j

r no. of attributes that equal 1 for object i but 0 for object j

s no. of attributes that equal 0 for object i and 1 for object j

t → no. of attributes that is zero for both objects i & j

for symmetric

$$d(i,j) = \frac{r+s}{q+r+s+t}$$

asymmetric

$$d(i,j) = \frac{r+s}{q+r+s}$$

Jaccard coefficient

similarly for asymmetric

$$sim_{Jacard}(i,j) = \frac{q}{q+r+s}$$

Name	Gender	Fever	Cough	Test-1	Test 2	Test3	Test	R-N
Jack	M	Y	N	P	N	N	N	
Mary	F	Y	N	P	N	P	N	
Jim	M	Y	P	N	N	N	N	

	Mary 1 (P)	Mary 0 (N)
Jack 1 (P)	1+1	0
0 (N)	1+1	1+1

	Mary 1	Mary 0
Jack 1	2	0
0	1	3

$d(\text{Jack}, \text{Mary}) = \frac{1}{3} = 0.33$

Cosine similarity

high value - similar
low value - less similar.

$\cos(i,j) = \frac{i \cdot j}{\|i\| \|j\|}$

R.No	age	income	student	credit rating	class buys computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle-aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle-aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle-aged	medium	no	excellent	yes
13	middle-aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

$$\Sigma - P_i \log_2 P_i$$

$$\text{info}(D) = -\frac{9}{14}\log\left(\frac{9}{14}\right) + -\frac{5}{14}\log\left(\frac{5}{14}\right)$$

$$= 0.94$$

$$\text{info}_{age} = \frac{5}{14} I(2,3) + \frac{4}{14}(0) + \frac{5}{14} I(3,2)$$

$$= \frac{5}{14}\left(-\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right)\right) + \frac{4}{14}\left(-\frac{4}{4}\log(1)\right) + \frac{5}{14}\left(-\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{3}{5}\right)\right)$$

$$Info_{income} = \frac{4}{14}(2,2) + \frac{6}{14}(4,2) + \frac{4}{14}(3,1)$$

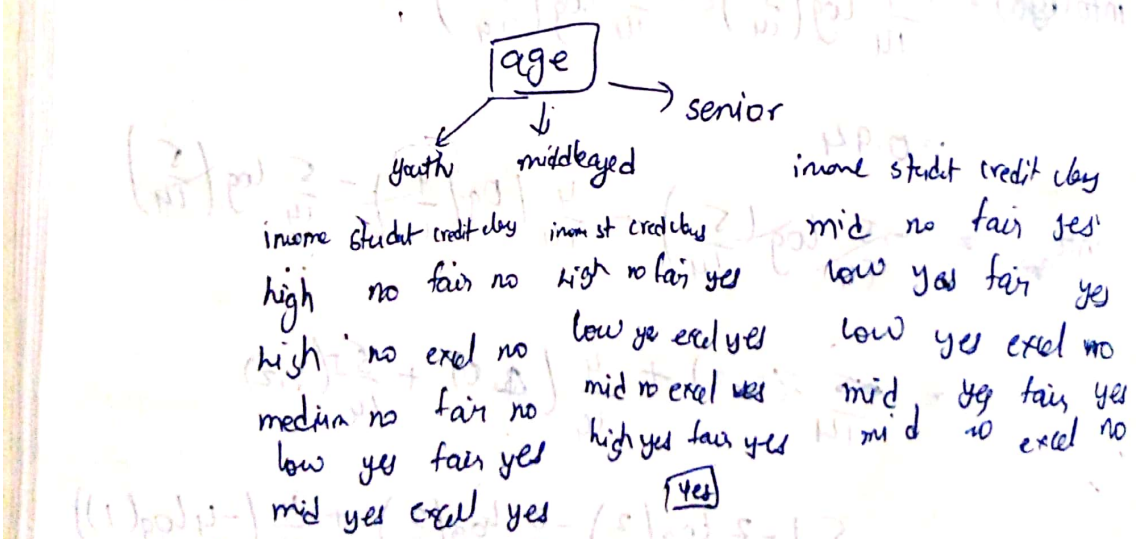
$$= \frac{4}{14}\left(-\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right)\right) + \frac{6}{14}\left(-\frac{4}{6}\log\left(\frac{4}{6}\right) - \frac{2}{6}\log\left(\frac{2}{6}\right)\right)$$

$$+ \frac{4}{14}\left(-\frac{3}{4}\log\left(\frac{3}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right)\right)$$

$$= 0.911$$

$$gain_{income} = 0.0292$$

$Info_{stud} = 0.7884$
$gain_{stud} = 0.1548$
$Info_{credit} = 0.8921$
$gain_{credit} = 0.0481$



For youth

$$Info(D) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.97$$

$$Info(income) = \frac{2}{5}(2,0) + \frac{2}{5}(1,1) + \frac{1}{5}(1,0)$$

$$= \frac{2}{5}\left(-\frac{2}{2}\log 1\right) + \frac{2}{5}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) + 0$$

$= \frac{2}{5} = 0.4$

$I_{train(income)} = 0.57$

$student = \frac{3}{5}(3,0) + \frac{2}{5}(2,0)$

$= 0$

$I_{train(st)} = 0.97$

$I_{credit} = \frac{3}{5}(2,1) + \frac{2}{5}(1,1)$

$= \frac{3}{5}(0.91) + \frac{2}{5}$

$= 0.946$

$Gain_{(credit)} = 0.97 - 0.95 = 0.02$

For senior

$I(D) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.97$

$I_{(income)} = \frac{3}{5}(2,1) + \frac{2}{5}(1,1)$

$= 0.946$

$G_{income} = 0.02$

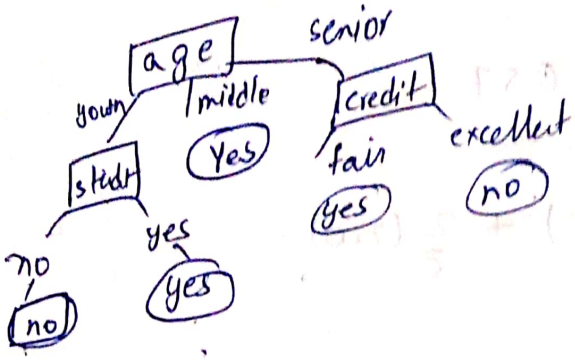
$I_{student} = \frac{3}{5}(2,1) + \frac{2}{5}(1,1)$

$= 0.946$

$G_{student} = 0.02$

$I_{credit} = \frac{3}{5}(3,0) + \frac{2}{5}(2,0)$

$G_{credit} = 0.97$



R: IF age = youth AND student = yes THEN buys_computer = yes

n_{covers} = #no. of tuples covered by R

$n_{correct}$ = no. of tuples correctly classified by R

$coverage(R) = n_{covers}(R) / |D|$

$accuracy(R) = n_{correct}(R) / n_{covers}(R)$

$$\text{Foil gain} = pos' \times \left(\log_2 \frac{pos'}{pos'+neg'} - \log_2 \frac{pos}{pos+neg} \right)$$

Outlook = S	$2\left(\log \frac{2}{5} - \log \frac{9}{14}\right) = -1.36$
Outlook = O	$4\left(\log 1 - \log \frac{9}{14}\right) = 2.54$
Outlook = R	$3\left(\log \frac{3}{5} - \log \frac{9}{14}\right) = -0.29$
T = H	$2\left(\log \frac{2}{4} - \log \frac{9}{14}\right) = -0.72$
T = M	$4\left(\log \frac{4}{6} - \log \frac{9}{14}\right) = 0.209$
T = C	$3\left(\log \frac{3}{4} - \log \frac{9}{14}\right) = 0.667$
H = H	
H = W	
W = W	
W = S	

$$\text{Foil prune} = \frac{pos - neg}{pos + neg}$$

	A	B	C	D,F	E
A	0	0.71	5.66	3.2	4.24
B	0.71	0	4.95	2.5	3.54
C	5.66	4.95	0	2.24	1.41
D,F	3.2	2.5	2.24	0	1.00
E	4.24	3.54	1.41	1.00	0

	A,B	C	D,F	E
A,B	0	4.95	2.5	3.54
C	4.95	0	2.24	1.41
D,F	2.5	2.24	0	1.00
E	3.54	1.41	1.00	0

	A,B	C	D,E,F
A,B	0	4.95	2.5
C	4.95	0	1.41
D,E,F	2.5	1.41	0

	A,B	C,D,E,F
A,B	0	2.5
C,D,E,F	2.5	0

Single linkage - min
complete linkage → max

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

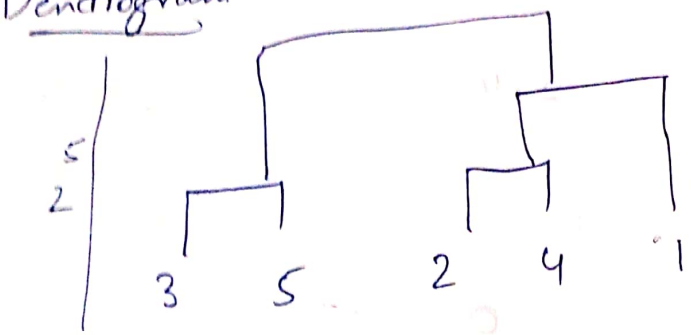
	3,5	1	2	4
3,5	0			
1	11	0		
2	10	9	0	
4	9	6	5	0

	3,5	1	2,4
3,5	0		
1	11	0	
2,4	10	9	0

	3,5	1,2,4
3,5	0	
1,2,4	11	0

1,2,3,4,5

Dendrogram

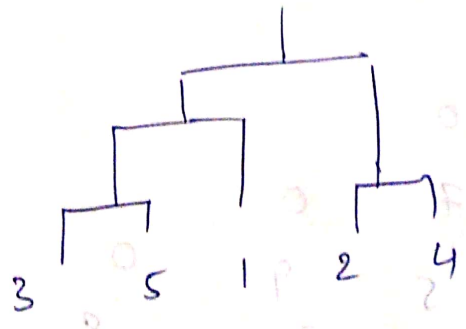


	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	10	11	2	8	0

1	2	3,5	4
1	0		
2	9	0	
3,5	(3)	7	0
4	6	5	8	0

1,3,5	2	4
1,3,5	0	
2	7	0
4	6	(5)	0

1,3,5	2,4
1,3,5	0
2,4	(6)	0



	A	B	C	D	E	F
A	0					
B	0.71	0				
C	5.66	4.95	0			
D	3.61	2.92	2.24	0		
E	4.24	3.54	1.41	1.0	0	
F	3.2	2.5	2.5	(0.5)	1.12	0

	A	B	C	D,F	E
A	0				
B	(0.71)	0			
C	5.66	4.95	0		
D,F	3.61	2.92	2.5	0	
E	4.24	3.54	1.41	(1.12)	0

	A,B	C	D,F	E
A,B	0			
C	5.66	0		
D,F	3.61	2.5	0	
E	4.24	1.41	(1.12)	0

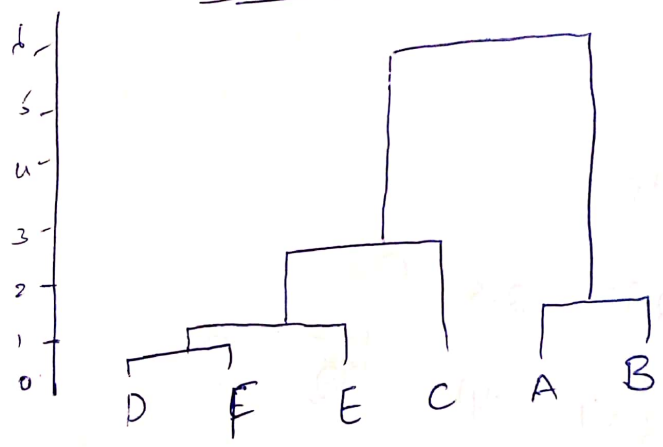
A, B C D, E, F

A, B	0		
C	5.66	0	
D, E, F	4.24	2.5	0

A, B C, D, E, F

A, B	0	
C, D, E, F	5.66	0

A, B, C, D, E, F



T_1 {A, B, C, D}

T_2 {A, D}

T_3 {A, E}

T_4 = {C, E}

Min support = 50%.

	Support
A	3
B	1
C	2
D	2
E	2

AC	1
AD	2
AE	1
CD	1
CE	1
DE	0

Closed freq {A} {C} {E} {AD} → no superset is in frequent set. → support not equal to any others

max pattern {C} {E} {AD}

TID	List of item IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Candidate set C_1

	Support
I_1	6
I_2	7
I_3	6
I_4	2
I_5	2

L_1

	Support
I_1	6
I_2	7
I_3	6
I_4	2
I_5	2

min support :

C_2

	support
I_1I_2	4
I_1I_3	4
I_1I_4	1
I_1I_5	2
I_2I_3	4
I_2I_4	2
I_2I_5	2
I_3I_4	0
I_3I_5	1
I_4I_5	0

L_2

I_1I_2	4
I_1I_3	4
I_1I_5	2
I_2I_3	4
I_2I_4	2
I_2I_5	2

C_3

	support
$I_1I_2I_3$	2
$I_1I_2I_5$	2
$I_1I_3I_5$	1
$I_2I_3I_4$	0
$I_2I_3I_5$	1
$I_2I_4I_5$	0

L_3

$I_1I_2I_3$	2
$I_1I_2I_5$	2

$I_1\ I_2\ I_3\ I_5$ | ✗

$I_1 I_2 I_3$ $I_1 I_2 I_5$

$I_1 \rightarrow I_2 I_3$ $\frac{2}{6}$

$I_2 \rightarrow I_1 I_3$ $\frac{2}{7}$

$I_3 \rightarrow I_1 I_2$ $\frac{2}{6}$

$I_1 I_2 \rightarrow I_3$ $\frac{2}{4}$ ✓

$I_2 I_3 \rightarrow I_1$ $\frac{2}{4}$ ✓

$I_1 I_3 \rightarrow I_2$ $\frac{2}{4}$ ✓

$I_1 \rightarrow I_2 I_5$ $\frac{2}{6}$

$I_2 \rightarrow I_1 I_5$ $\frac{2}{7}$

$I_5 \rightarrow I_1 I_2$ $\frac{2}{2}$ ✓

$I_1 I_2 \rightarrow I_5$ $\frac{2}{4}$ ✓

$I_2 I_5 \rightarrow I_1$ $\frac{2}{2}$ ✓

$I_1 I_5 \rightarrow I_2$ $\frac{2}{2}$ ✓

Tid	
10	A C D
20	B C E
30	A B C E
40	B E

Min support: 2

A	2
B	3
C	3
D	1 ✗
E	3

A	2
B	3
C	3
E	3

AB
AC
AE
BC
BE
CE

AB	1 ✗
AC	2
AE	1 ✗
BC	2
BE	3
CE	2

BCE → 2

$B \rightarrow C\ E$ $\frac{2}{3}$

$C \rightarrow B\ E$ $\frac{2}{3}$

$E \rightarrow B\ C$ $\frac{2}{3}$

$BC \rightarrow E$ $\frac{2}{2}$ $\frac{2}{3}$

$BE \rightarrow C$

$B \rightarrow C$ $\frac{2}{3}$

$C \rightarrow B$ $\frac{2}{3}$

$B \rightarrow E$ $\frac{3}{3}$

$E \rightarrow B$ $\frac{3}{3}$

FP Growth :-

Min support: 2

T	Items
T_{100}	I_1, I_2, I_5
T_{200}	I_2, I_4
T_{300}	I_2, I_3
T_{400}	I_1, I_2, I_4
T_{500}	I_1, I_3
T_{600}	I_2, I_3
T_{700}	I_1, I_3
T_{800}	I_1, I_2, I_3, I_5
T_{900}	I_1, I_2, I_3

Step1: Generate Itemset & their frequency (support)

I_1	6
I_2	7
I_3	6
I_4	2
I_5	2

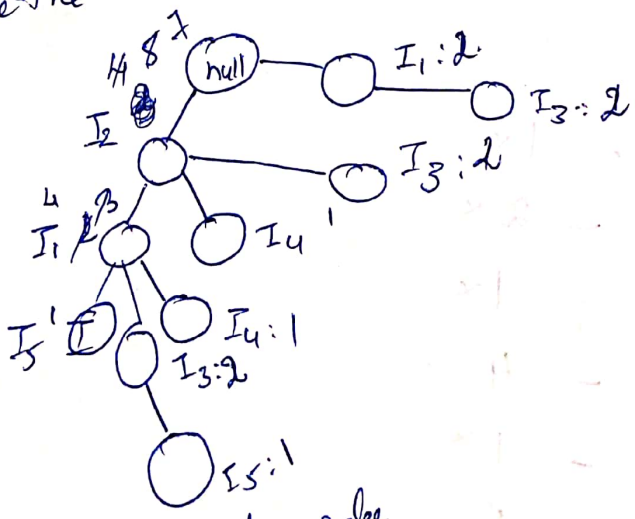
Step2: Sort in descending order

I_2	7
I_1	6
I_3	6
I_4	2
I_5	2

Step3: Arrange the transaction as per the order in step 2

T	Items
T_{100}	$I_2 I_1 I_5$
T_{200}	$I_2 I_4$
T_{300}	$I_2 I_3$
T_{400}	I_2, I_1, I_4
T_{500}	$I_1 I_3$
T_{600}	~~I_2~~ I_3
T_{700}	$I_1 I_3$
T_{800}	$I_2 I_1 I_3 I_5$
T_{900}	$I_2 I_1 I_3$

Step 4: Draw the tree



Step5: Sort Items in ascendy order

Items	Conditional Pattern Base
I_5	$\{I_2 I_1 : 1\}, \{I_2 I_1 I_3 : 1\}$
I_4	$\{I_2 : 1\}, \{I_2 I_1 : 1\}$
I_3	$\{I_2 : 2\}, \{I_2 I_1 : 2\}, \{I_1 : 2\}$
I_1	$\{I_2 : 4\}$, ~~$\{null : 2\}$~~
I_2	~~$\{null : 1\}$~~

Items	Conditional Pattern Base	Condition F.P
I_5	$\{I_2 I_1 : 1\}\{I_2 I_1 I_3 : 1\}$	$\langle I_2 : 2, I_1 : 2 \rangle$
I_4	$\{I_2 I_1 : 1\}\{I_2 : 1\}$	$\{I_2 : 2\}$
I_3	$\{I_2 I_1 : 2\}\{I_2 : 2\}\{I_1 : 2\}$	$\{I_2 : 4, I_1 : 2\}$
I_1	$\{I_2 : 4\}$	$\{I_2 : 4\}$

T_{100} MONKEY.

T_{200} DONKEY

T_{300} MAKE.

T_{400} MUCKY.

T_{500} COOKIE

min sup: 3

Step 1:

A - 1 x

C - 2 x

D - 1 x

E - 4 ✓

I - 1 x

K - 5 ✓

M - 3

N - 2 x

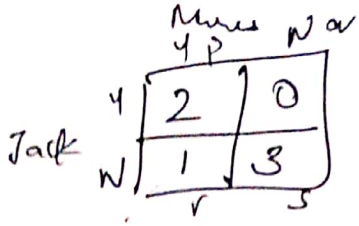
O - 4 ✓

U - 1 x

Y - 3

Step 2:

K	5
E	4
O	4
M	3
Y	3



Step 3:

T_{100} KEOMY

T_{200} KEOY

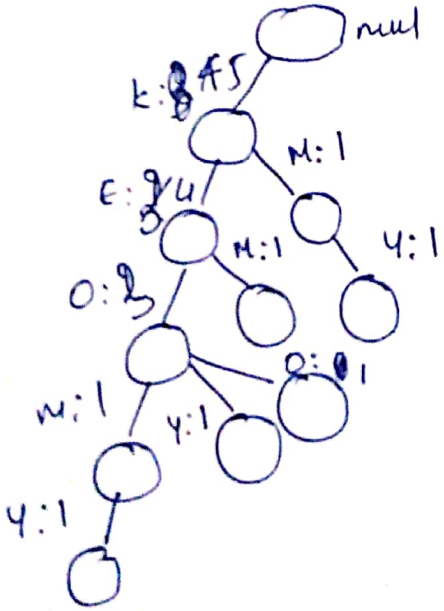
T_{300} KEM

T_{400} KMY

T_{500} KEO

Symmetri $\frac{0+1}{0+1+2+3}$

or $d = \frac{0+1}{0+1+2}$



Si $L = 1 - d$

$\frac{2}{0+1+2}$

Y {KEOM:1, KEO:1, KM:1} <K:3>

M {K:1, KE:1, KEO:1} <K:3>

O {KE:3, KEO:1}

E {K:4} <K:4>

K

KY - 3 KE:4

KM - 3