DMBI Nov-17

(8)

Transactions	Items
T1	Bread, jelly, milk
T2	Butter, jelly, juice.
T3	Bread, Butter, jelly, juice
T4	Butter, juice

Support = 50% Confidence = 70%

$= \frac{50}{100} \times 4 = 2$

S1 :-> Find support count of each Itemset -> C1

Items	Frequency [count]
Bread	2
jelly	3
milk	1
Butter	3
Juice	3

S2 :-> Compare min support count with each itemsets support count (L1).

Items	Frequency [count]
Bread	2
jelly	3
Butter	3
juice	3

S3 :-> Generate pair from L1 to generate support count [C1] :->

Items	Frequency [count]
{Bread, jelly}	2
{Bread, Butter}	1
{Bread, juice}	1

{jelly, Butter}	2
{jelly, juice}	2
{Butter, juice}	3

S4 :-> Compare min L_2

Items	Frequency [count]
{Bread, jelly}	2
{jelly, Butter}	2
{jelly, juice}	2
{Butter, juice}	3

S5 :-> Generate pair from L_2, C_3.

Items	Frequency [count]
{Bread, jelly, Butter}	1
{Bread, jelly, juice}	1
{jelly, Butter, juice}	2
~~{jelly}~~	

S6 :-> Compare min L_3

Items	Frequency [count
{jelly, Butter, juice}	2

S7 :-> L_3 has only 1 item pair, there is no

Association Item Rule.	Support	Confidence	
Jelly ^ Butter => juice	2	1	100%
Jelly ^ juice => Butter	2	1	100%
Butter ^ juice => jelly	2	0.66	66%
juice => jelly ^ Butter	2	0.66	66%
Butter => jelly ^ juice	2	0.66	66%
jelly => Butter ^ juice	2	0.66	66%

(2)

$$\text{Confidence} = \frac{\text{Support}}{\text{no. of time left item occurs}}$$

$\frac{2}{3} \times 0.$

Since Confidence = 70%

Compare with min-confidence = 70%

rules	Support	confidence.
jelly ^ Butter => juice	2	100
jelly ^ juice => Butter	2	100%

Hence Final Association rules are

jelly ^ Butter => juice

jelly ^ juice => Butter

Naive Bayes

K-Means - 1A Single

$K = \{2, 3, 4, 10, 11, 12, 20, 25, 30\}$

$k = 2$ $\quad m_1 = 4$ $\quad m_2 = 12$

$K_1 = \{2, 3, 4\}$

$K_2 = \{10, 11, 12, 20, 25, 30\}$

$m_1 = \frac{9}{3} = 3$, $m_2 = \frac{108}{6} = 18$

$K_1 = \{2, 3, 4, 10\}$ $\quad K_2 = \{11, 12, 20, 25, 30\}$

$m_1 = \frac{19}{4} = 4.75 \rightarrow 5$ | $m_2 = 19.6 \rightarrow 20$

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad K_2 = \{20, 25, 30\}$

$m_1 = 7$ $\quad m_2 = 25$

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad K_2 = \{20, 25, 30\}$

$m_1 = 7$ $\quad m_2 = 25$

Naive Bayes

Outlook	Temp	Humidity	Windy	PlayGolf
Rainy	H	H	F	N
R	H	H	T	N
O	H	H	F	Y
S	M	H	F	Y
S	C	N	F	Y
O	C	N	T	N
R	C	N	T	Y
S	M	H	F	N
R	M	N	F	Y
O	M	N	T	Y
O	M	H	T	Y
S	H	N	F	Y
	M	H	T	N

PlayGolf

Play	Y	N	P(Y)	P(N)
	9	5	9/14	5/14

Outlook

	Y	N	P(Y)	P(N)
R	2	3	2/9	3/5
O	4	0	4/9	0/5
S	3	2	3/9	2/5
	9	5	9/9	5/5

Temp

	Y	N	P(Y)	P(N)
H	2	3	2/9	3/5
M	4	1	4/9	1/5
C	3	1	3/9	1/5
	9	5	9/9	5/5

Humidity

	Y	N	P(Y)	P(N)
H	3	4	3/9	4/5
N	6	1	6/9	1/5

Windy

	Y	N	P(Y)	P(N)
T	3	3	3/9	3/5
F	6	2	6/9	2/5

If the whether is sunny, temp is hot, humidity is high, & there is no wind then can we play golf.

Yes.

$$P(c/x) = \frac{\Sigma P(x/c) \times P(c)}{P(x)}$$

c → class to predict

x = Sunny, hot, Normal, False..

c = Yes.

$$\approx P(\text{Sunny}/\text{Yes}) \times P(\text{hot}/\text{Yes}) \times P(\text{high}/\text{Yes}) \times P(F/\text{Yes}) \times P(\text{Yes})$$

$$\approx \frac{3}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9} \times \frac{9}{14}$$

$$\approx 0.02116$$

* Regression Analysis (If exp lo years (4) how much sal)

x (yrs) Experience	y salary (in $1000)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i-\bar{x})\cdot(y_i-\bar{y})$	$(x_i-\bar{x})^2$
3	30	-6.1	-25.4	154.94	37.21
8	57	-1.1	1.6	-1.76	1.21
9	64	-0.1	8.6	-0.86	0.01
13	72	3.9	16.6	64.74	15.21
3	36	-6.1	-19.4	118.34	37.21
6	43	-3.1	-12.4	38.44	9.61
11	59	1.9	3.6	6.84	3.61
21	90	11.9	34.6	411.74	141.61
1	20	-8.1	-35.4	286.74	65.61
16	33	6.9	27.6	190.44	47.61
				1269.6	358.9

(m)

$$w_1 = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10} (x_i - \bar{x})^2} \rightarrow m \text{ (slope)}$$

(c)

$$w_0 = \bar{y} - w_1 \bar{x}$$

$$\bar{x} = \frac{3+8+9+13+3+6+11+21+1+16}{10} = 9.1$$

$$\bar{y} = \frac{30+57+64+72+36+43+59+90+20+33}{10} = 55.4$$

$w_1 = 3.54$ $\quad w_0 = 23.18$

$y = w_0 + w_1 x \quad (y = mx + c)$

$= 23.18 + 3.54x$

If exp - 10 years - x

$= 23.18 + 3.54(10) =$

14 year of exp is 14 then salary will be :-

$$y = 23.18 + 3.54(14) = 72.74$$

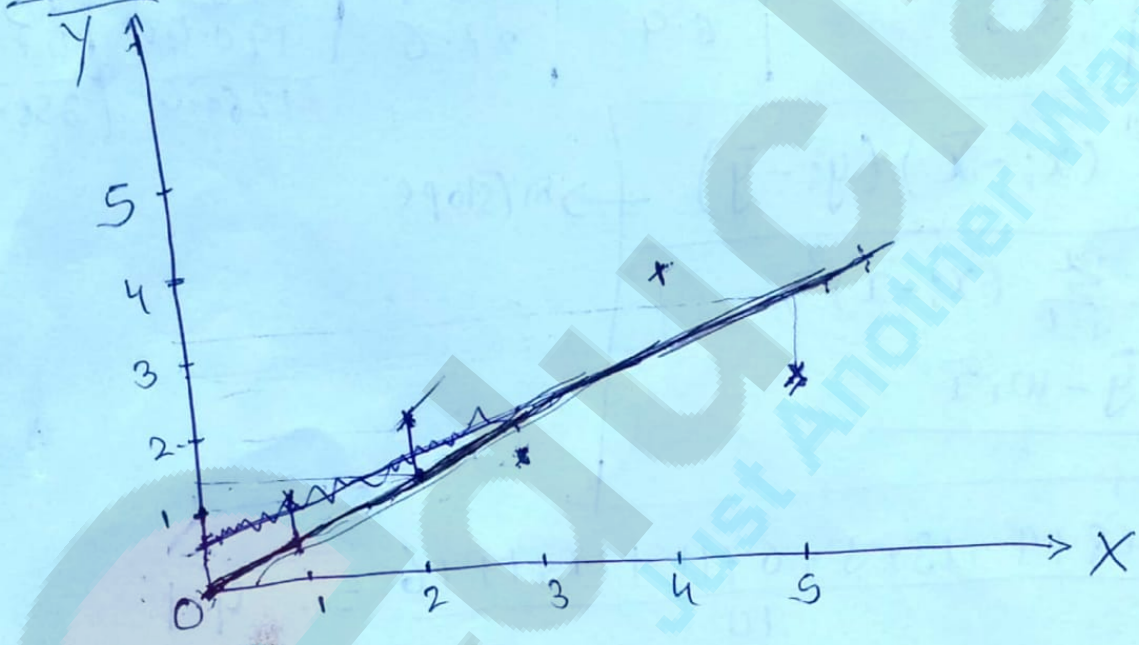
★ Linear. Reg

Given in ques.

X	Y	y'	y − y'	$(y-\bar{y}')^2$
1.00	1.00	0.75	0.25	0.0625
2.00	2.00	1.2	0.8	0.64
3.00	1.30	2.0	=0.7	0.49
4.00	3.75	2.9	0.85	0.722
5.00	2.25	3.2	−0.95	0.9025

1.30

Graph



X, Y y' y − y

1

$y = mx + c$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{3.75 - 1}{4 - 1} = \frac{2.25}{3} = 0.9166$$

Assume For $x = 1$

$= mx + c$

$y = 0.9166 \times 1 + 0$

$y = 0.9166$

(5)

0.022

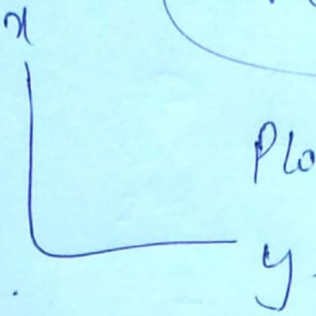
Agglomerative [Single Link]. Min.

[Complete Link (max)]

[Average Link → average]

	X	Y
P_1	0.40	0.53
P_2	0.22	0.38
P_3	0.35	0.32
P_4	0.26	0.19
P_5	0.08	0.41
P_6	0.45	0.30

Distance = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$



Plot graph.

(P_1, P_2)

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	0.23	0.21	0.25	0.3417	0.2353
P_2	0.234	0			0.1431	
P_3	0.22	0.15	0			
P_4	0.37	0.20	0.15	0		
P_5	0.3417	0.14 (~~0.1431~~)	0.28	0.29	0	
P_6	0.23	0.25	0.11	0.22	0.39	0

0.11 → min value

P_3, P_6 → 1 Block

	P_1	P_2	P_3	P_4	P_5
P_1	0				
P_2	0.24	0			
P_3, P_6	0.22	0.15	0		
P_4	0.36	0.20	0.15	0	
P_5	0.34	0.14	0.28	0.29	0

Calculate the distance P_1, P_3 & P_1, P_6 and the min value will be written.