1) $n = 300$

$\sigma^2 = 225$ $\quad \sigma = \sqrt{225} = 15$

$\bar{x} = 54$ $\quad Z(\alpha/2)$ at 95%

$\frac{7.5}{15}$ $2$

$\Rightarrow \sigma(\alpha/2)$ at 95% $= 1.96$

$M \cdot E(M) = Z_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow 1.96 \cdot \frac{15}{\sqrt{300}}$

$\Rightarrow 1.6974$

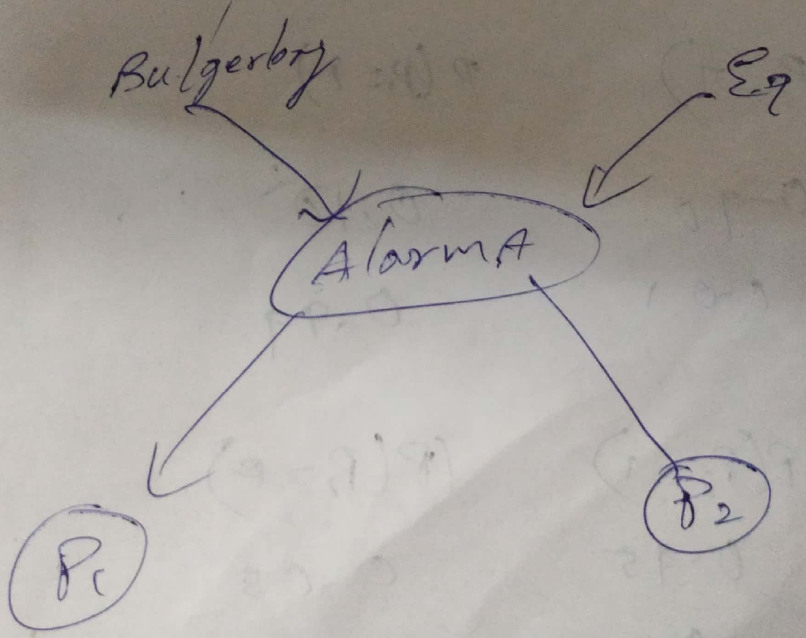
$(54 - 1.6974,\ 54 + 1.6974)$

$\Rightarrow (52.3026,\ 55.6974)$

---

M-E

## bayesian belief network



Probability of B & E

P(B=T) = 0.999

P(B=F) = 0.992

P(E=T) = 0.998

P(E=F) = 0.994

Probablity of Alarm

| B | E | P(A=T) | P(A=F) |
|---|---|---|---|
| T | T | 0.95 | 0.05 |
| T | F | 0.99 | 0.01 |
| F | T | 0.93 | 0.07 |
| F | F | 0.001 | 0.999 |

probablity of person ($P_1$)

| A | $P(P_1=T)$ | $P(P_1=F)$ |
|---|---|---|
| T | 0.90 | 0.10 |
| F | 0.01 | 0.99 |

Person ($P_2$)

| A | $P(P_2=T)$ | $P(P_2=F)$ |
|---|---|---|
| T | 0.95 | 0.05 |
| F | 0.02 | 0.98 |

Scanned with OKEN Scanner

Eg. find probability of $P_1$ is T, $P_2$ is T, A is T, B is F, E is F

$P(P_1, P_2, A, \sim B, \sim E)$

$\Rightarrow P(P_1|A)\ P(P_2|A)\ P(A|\sim B \sim E)\ P(\sim B)\ P(\sim E)$

$\Rightarrow 0.90 \times 0.95 \times 0.001 \times 0.992 \times 0.994$

$\Rightarrow 0.00084$

min-sup = 40%

min-conf = 70%

# Apriori Algorithm

| Tid | |
|---|---|
| $T_1$ | Br, Bu, m |
| $T_2$ | Br, Bu |
| $T_3$ | Bi, c, d |
| $T_4$ | Br, Bu, m, d |
| $T_5$ | Bi, d |

| | support | min_support |
|---|---|---|
| Br | 3 | 3/5 = 60% |
| Bu | 3 | 60% |
| m | 2 | 40% |
| Bi | 2 | 40% |
| ~~c~~ | 1 | 20% (x) |
| d | 3 | 60% |

{Br, Bu, m, Bi, d}

(Br, Bu) (Br, m) (Br, Bi) (Br, d)

(Bu, m) (Bu, Bi) (Bu, d) (m, Bi) (m, d)

(Bi, d)

Scanned with OKEN Scanner

| | support | min-support | |
|---|---|---|---|
| (Br, Bu) | 3 | 3/4 0.60% | |
| (Br, m) | 2 | 40% | |
| ~~(Br, Bi)~~ | ~~0~~ | ~~0~~ | (x) |
| ~~(Br, d)~~ | ~~1~~ | ~~20%~~ | (x) |
| (Bu, m) | 2 | 40% | |
| ~~(Bu, Bi)~~ | ~~0~~ | ~~0~~ | |
| ~~(Bu, d)~~ | ~~1~~ | 20% | |
| ~~(m, Bi)~~ | ~~0~~ | 0 | |
| ~~(m, d)~~ | ~~1~~ | 20% | |
| (Bi, d) | 2 | 40% | |

(Br, Bu) (Br, m) (Bu, m) (Bi, d)

~~(Br, Bu, m)~~ ~~(Br,~~

| | Support | min sup |
|---|---|---|
| (Br, Bv, m) | 2 | 40% |
| (Br, Bv, Bi) | 0 | 0 |
| Br, Bv, d | 1 | 20% |
| Br, m, Bi | 0 | |
| Br, m, d | 1 | 20% |
| Br, Bi, d | 0 | |
| Bv, m, Bi | 0 | |
| Bv, m, d | 0 | |
| Bv, Bi, d | 0 | |
| (m, Bi, d) | 0 | |

(Br, Bv, m)  min_confidence = 70%

| | Support | confidence |
|---|---|---|
| $(Br \wedge Bv) \rightarrow m$ | 2 | 66% |
| $(Br \wedge m) \rightarrow Bv$ | 2 | 100% |
| $(Bv \wedge m) \rightarrow Br$ | 2 | 100% |
| $Br \rightarrow (Bv \wedge m)$ | 2 | 66% |
| $Bv \rightarrow (Br \wedge m)$ | 2 | 66% |
| $m \rightarrow (Br \wedge Bv)$ | 2 | 100% |

Scanned with OKEN Scanner

$$confidence = \frac{Support(A \cup B)}{Support(A)}$$

$\underbrace{(2,3)}_{A} \rightarrow \underbrace{5}_{B}$

$~~2~~$

$\underbrace{2}_{A} \rightarrow \underbrace{(3,5)}_{B}$

$$\Rightarrow \frac{S(\cancel{Br, Bu, m})}{S(\cancel{Br, B})} \Rightarrow \frac{2}{3} \Rightarrow 66\%$$

$$\Rightarrow \frac{S(Br, Bm, BU)}{S(Br, m)} = \frac{2}{2} \Rightarrow 100\%$$

$$\Rightarrow \frac{S(BU, m, Br)}{S(BU, m)} = \frac{2}{2} = 100\%$$

$$\Rightarrow \frac{S(Br, BU, m)}{\cancel{S(BU, m)}\ S(Br)} = \frac{2}{3} \Rightarrow 66\%$$

$$\Rightarrow \frac{S(BU, Br, m)}{S(BU)} = \frac{2}{3} \Rightarrow 66\%$$

$$\Rightarrow \frac{S(m, Br, BU)}{S(m)} = \frac{2}{2} \Rightarrow 100\%$$

$(Br \wedge m) \rightarrow BU$ , $(BU \wedge m) \rightarrow Br$ , $m \rightarrow (Br \wedge BU)$

Strong associations.

min-sup = 30%

| | |
|---|---|
| 1 | ~~Pn~~, h, r |
| 2 | pl, e, s |
| 3 | Pn, r, ch, s |
| 4 | pl, cl, e |
| 5 | r, Pi, sb, Pn |
| 6 | m, ch, sk |

| | support | min_support |
|---|---|---|
| Pn | 3 | 50% |
| ~~h~~ | ~~1~~ | ~~16%~~ |
| r | 3 | 50% |
| pl | 2 | 33% |
| e | 2 | 33% |
| s | 2 | 33% |
| ch | 2 | 33% |
| ~~cl~~ | ~~1~~ | ~~16%~~ |
| ~~Pi~~ | ~~1~~ | ~~16%~~ |
| ~~sb~~ | ~~1~~ | ~~16%~~ |
| ~~m~~ | ~~1~~ | ~~16%~~ |
| ~~sk~~ | ~~1~~ | ~~16%~~ |

(Pn, r, pl, e, s, ch)

(Pn, r) (Pn, pl) (Pn, e) (Pn, s) (Pn, ch)

(r, pl) (r, e) (r, s) (r, ch)

(pl, e) (pl, s) (pl, ch) (e, s) (e, ch) (s, ch)

| | | |
|---|---|---|
| Pn, r | 3 | 50% |
| Pn, pl | 0 | |
| Pn, e | 0 | |
| Pn, s | 1 | 16% |
| Pn, ch | 1 | 16% |
| r, pl | 0 | |
| r, e | 0 | |
| r, s | 1 | 16% |
| r, ch | 1 | 16% |
| pl, e | 2 | ~~50%~~ 33% |
| pl, s | 1 | 16% |
| pl, ch | 0 | |
| e, s | 1 | 16% |
| e, ch | 0 | |
| s, ch | 1 | 16% |

(pn, r) (pd, e)

(pn, r, pl) (pn, r, e)

| | |
|---|---|
| ~~pn, r, pl~~ | 0 |
| ~~pn, r, e~~ | 0 |

Ch

El

P

# FP Growth Algorithm

FP → frequent pattern.

- is an efficient and scalable method for mining the complete set of FP using a tree structure for storing information about FP called FP tree.

Eg :- min_support = 30%.

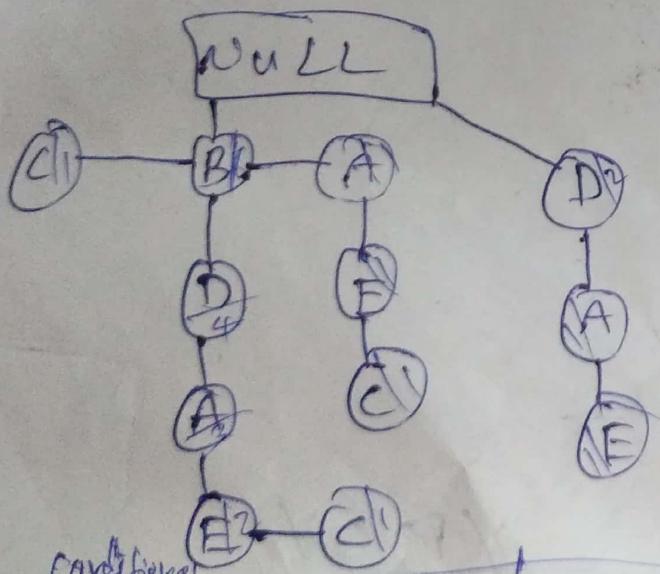
| Tid | items |
|---|---|
| 1 | E, A, D, B |
| 2 | D, A, E, C, B |
| 3 | C, A, B, E |
| 4 | B, A, D |
| 5 | D |
| 6 | D, B |
| 7 | A, D, E |
| 8 | B, C |

| | support | | priority |
|---|---|---|---|
| A | ~~4~~ 5 | 5/8 ⇒ 62% | 3 |
| B | 6 | | 1 |
| C | 3 | 37% | 5 |
| D | 6 | | 2 |
| E | 4 | | 4 |

## Priority order

(B, D, A, E, C)

| Tid | | Items | Ordered items |
|---|---|---|---|
| A | 1 | E, A D, B | B, D, A, E |
| B | 2 | D, A, E, C, B | B, D, A, E, C |
| C | 3 | C A, B, E | B, A, E, C |
| D | 4 | B, A, D | B, D, A |
| E | 5 | D | D |
| | 6 | D, B | B, D |
| | 7 | A, D, E | D, A, E |
| | 8 | B, C | B, C |




B – 1, 2, 3, 4,
D – 1, 2, 3, 4
A – 1, 2, 3,
E – 1, 2
C – 1
A – 1,
E – 1
C – 1
D – 1, 2
A – 1
E – 1
C – 1

Conditional

| | | | | |
|---|---|---|---|---|
| A | ~~conditional~~ {B, 1} {B, D, 2} {D, 1} | {Ø | 7 | 9 |
| B | {G} | {Ø} | | |
| C | {B, 1} {A, E, 1} {B, D, A, E, 1} | Ø | | |
| D | {B, 4} {2} | Ø | | |

# K-Medoid clustering or PAM (Partition Around medoids)

- Apply K-Medoid clustering algo to form two clusters
- use Manhattan distance to find the between data point & medoid.

step-1

| i | x | y | $c_1$ | $c_2$ | cluster |
|---|---|---|---|---|---|
| $x_1$ | 2 | 6 | 3 | 7 | $c_1$ |
| $x_2$ | 3 | 4 | 0 | 4 | $c_1$ |
| $x_3$ | 3 | 8 | 4 | 8 | $c_1$ |
| $x_4$ | 4 | 7 | 4 | 6 | $c_1$ |
| $x_5$ | 6 | 2 | 5 | 3 | $c_2$ |
| $x_6$ | 6 | 4 | 3 | 1 | $c_2$ |
| $x_7$ | 7 | 3 | 5 | 1 | $c_2$ |
| $x_8$ | 7 | 4 | 4 | 0 | $c_2$ |
| $x_9$ | 8 | 5 | 6 | 2 | $c_2$ |
| $x_{10}$ | 7 | 6 | 6 | 2 | $c_2$ |

selecting two medoids

$c_1 = (3,4)$

$c_2 = (7,4)$

$|x_1 - x_2| + |y_1 - y_2|$

$\Rightarrow |2-3| + |6-4|$

$\Rightarrow 1 + 2$

$\Rightarrow 3$

with $c_2$

$|x_1 - x_2| + |y_1 - y_2|$

$\Rightarrow |2-7| + |6-4|$

$\Rightarrow 7$

Step-2

Clusters are

$C1 : \{ (2,6)\ (3,4)\ (3,8),\ (4,7) \}$

$C_2 : \{ (6,2)\ (6,4)\ (7,3)\ (7,4)\ (8,5)\ (7,6) \}$

* Now calcu ~~dista~~ cost with respecto to medoid

$cost(c, x) = \Sigma_i |c_i - x_i|$

Total cost

$|3-2| + |4-6|$
$\Rightarrow 1+2 \Rightarrow 3$

$\Rightarrow \{ cost[(3,4)\ (2,6)] + cost[(3,4)\ (3,8)] + cost[(3,4)(4,7)]$

$+ cost[(7,4)(6,2)] + cost[(7,4)\ (6,4)] + cost[(7,4)(7,3)]$

$+ cost[(7,4)\ (8,5)] + cost[(7,4)\ (7,6)]$

$\Rightarrow 3 + 4 + 4 + 2 + 3 + 1 + 1 + 2 = 20$

step-3

* Randomly select a non medoid point
* ~~find clust~~ recalculate cost.

$c_1 = (3,4)$ & $c_2 = (7,4)$

$O = (7,3)$

swap $c_2$ with $O$

New medoids

$c_1 = (3,4)$ & $O = (7,3)$

| i | x | y | $c_1$ | O | cluster |
|---|---|---|---|---|---|
| $x_1$ | 2 | 6 | 3 | 8 | 9 |
| $x_2$ | 3 | 4 | 0 | 5 | $c_1$ |
| $x_3$ | 3 | 8 | 4 | 9 | $c_1$ |
| $x_4$ | 4 | 7 | 4 | 7 | $c_1$ |
| $x_5$ | 6 | 2 | 5 | 2 | O |
| $x_6$ | 6 | 4 | 3 | 2 | O |
| $x_7$ | 7 | 3 | 5 | 0 | O |
| $x_8$ | 7 | 4 | 4 | 1 | O |
| $x_9$ | 8 | 5 | 6 | 3 | O |
| $x_{10}$ | 7 | 6 | 6 | 3 | O |

New clusters are

C1 :- {(2,6) (3,4) (3,8) (4,7)}

C2 :- {(6,2) (6,4) (7,3) (7,4) (8,5) (7,6)}

calcu cost

Total cost = {cost{(3,4) (2,6)} –

C

⇒ 3+4+4+2+2+1+3+3=22

S = current cost – previous cost.

⇒ 22 – 20 ⇒ 2 > 0

• Hence swapping C2 with O is not a good idea

final medoids : C1 = (3,4) C2 = (7,4)

clusters are

C1 = {

C2 = {

# K-means

means 2 clusters

12, 24, 20, 42, 15, 14, 11, 9 & $K = 2$

$K = \{9, 11, 12, 14, 15, 20, 24, 42\}$

10.5

a) $m_1 = 12$ $m_2 = 20$

Step1:-

$K_1 = \{9, 11, 12, 14, 15\}$

$K_2 = \{20, 24, 42\}$

$m_1 = \frac{9+11+12+14+15}{5}$ $m_2 = \frac{20+24+42}{3}$

$m_1 = 12.2$ $m_2 = 28.6$

Step2:-

$K_1 = \{9, 11, 12, 14, 15, 20\}$ $K_2 = \{24, 42\}$

$m_1 = 13.5$ $m_2 = 33$

Step3:-

$K_1 \{9, 11, 12, 14, 15, 20\}$ $K_2 = \{24, 42\}$

$m_1 = 13.5$ $m_2 = 33$

we are getting same mean, so the new clusters are:

$K_1 = \{9, 11, 12, 14, 15, 20\}$ $K_2 = \{24, 42\}$

# K-NN

Maths = 6, CS = 8 & K = 3

3P OF

| | Maths | CS | Result |
| --- | --- | --- | --- |
| 1) | 4 | 3 | F |
| 2) | 6 | 7 | P |
| 3) | 7 | 8 | P |
| 4) | 5 | 5 | F |
| 5) | 8 | 8 | P |

Euclidean distance

$\sqrt{(x_{01} - x_{A0})^2 + (x_{02} - x_{A2})^2}$

$d_1 = \sqrt{(6-4)^2 + (8-3)^2} \Rightarrow 5.38$

$d_2 = \sqrt{(6-6)^2 + (8-7)^2} \Rightarrow 1$

$d_3 = \sqrt{(6-7)^2 + (8-8)^2} \Rightarrow 1$

$d_4 = \sqrt{(6-5)^2 + (8-5)^2} = 3.16$

$d_5 = \sqrt{(6-8)^2 + (8-8)^2} = 2$

2, 3, 5

final = Pass

A.D = 3 S = 7 A = 3 → 3 neibours

| A.D | S.S | class |
|---|---|---|
| 7 | 7 | Bad |
| 7 | 4 | Bad |
| 3 | 4 | Good |
| 1 | 4 | Good |

$d_1 = \sqrt{(3-7)^2 + (7-7)^2} = \sqrt{16+0} \Rightarrow 4$

$d_2 = \sqrt{(3-7)^2 + (7-4)^2} \Rightarrow \sqrt{16+9} \Rightarrow 5$

$d_3 = \sqrt{(3-3)^2 + (7-4)^2} \Rightarrow \sqrt{9} \Rightarrow 3$

$d_4 = \sqrt{(3-1)^2 + (7-4)^2} \Rightarrow \sqrt{13} \Rightarrow 3.6056$

| Good | Bad |
|---|---|
| 2 | ~~0~~ 1 |

final = Good ✓

⇒ outlook = rainy, temp = cool, humidity = high,

windy = true, play = ?

P(play = yes) = 9/14 = 0.6429

P(play = no) = 5/14 = 0.3571

| Outlook | Y | N |
|---|---|---|
| sunny | 2/9 | 3/5 |
| overcast | 4/9 | 0 |
| rainy | 3/9 | 2/5 |

| Humidity | Y | N |
|---|---|---|
| high | 3/9 | 4/5 |
| normal | 6/9 | 1/5 |

| temp | Y | N |
|---|---|---|
| hot | 2/9 | 2/5 |
| mild | 4/9 | 2/5 |
| cold | 3/9 | 1/5 |

| windy | Y | N |
|---|---|---|
| True | 3/9 | 3/5 |
| False | 6/9 | 2/5 |

$V_{NB}(yes) = P(yes)\, P(\cancel{sunny}\ rainy|yes)\, P(cool|yes)$

$P(high|yes)\ P(true|yes)$

$\Rightarrow 0.6429 \times 0.3333 \times 0.3333 \times 0.3333 \times 0.3$

$\Rightarrow 0.8572$

$V_{NB}(no) = P(no)\ P(rainy/no)\ P(cool/no)\ P(high/no)$
$P(true/no)$

$\Rightarrow 0.3571 \times 0.4 \times 0.2 \times 0.8 \times 0.6$

$\Rightarrow 0.0137$

yes is more than no

$\rightarrow$ so play = yes

Normalizing

$$V_{NB}(yes) = \frac{V_{NB}(yes)}{V_{NB}(yes) + V_{NB}(no)} \Rightarrow \frac{0.6429}{0.6429 + 0.3571}$$

$\Rightarrow 0.6429$

$$V_{NB}(no) = \frac{V_{NB}(no)}{(yes) + (no)} = \frac{0.3571}{0.6429 + 0.3571}$$

$\Rightarrow 0.3571$

for our checking

$0.6429 + 0.3571$

$\Rightarrow 1$ ✓