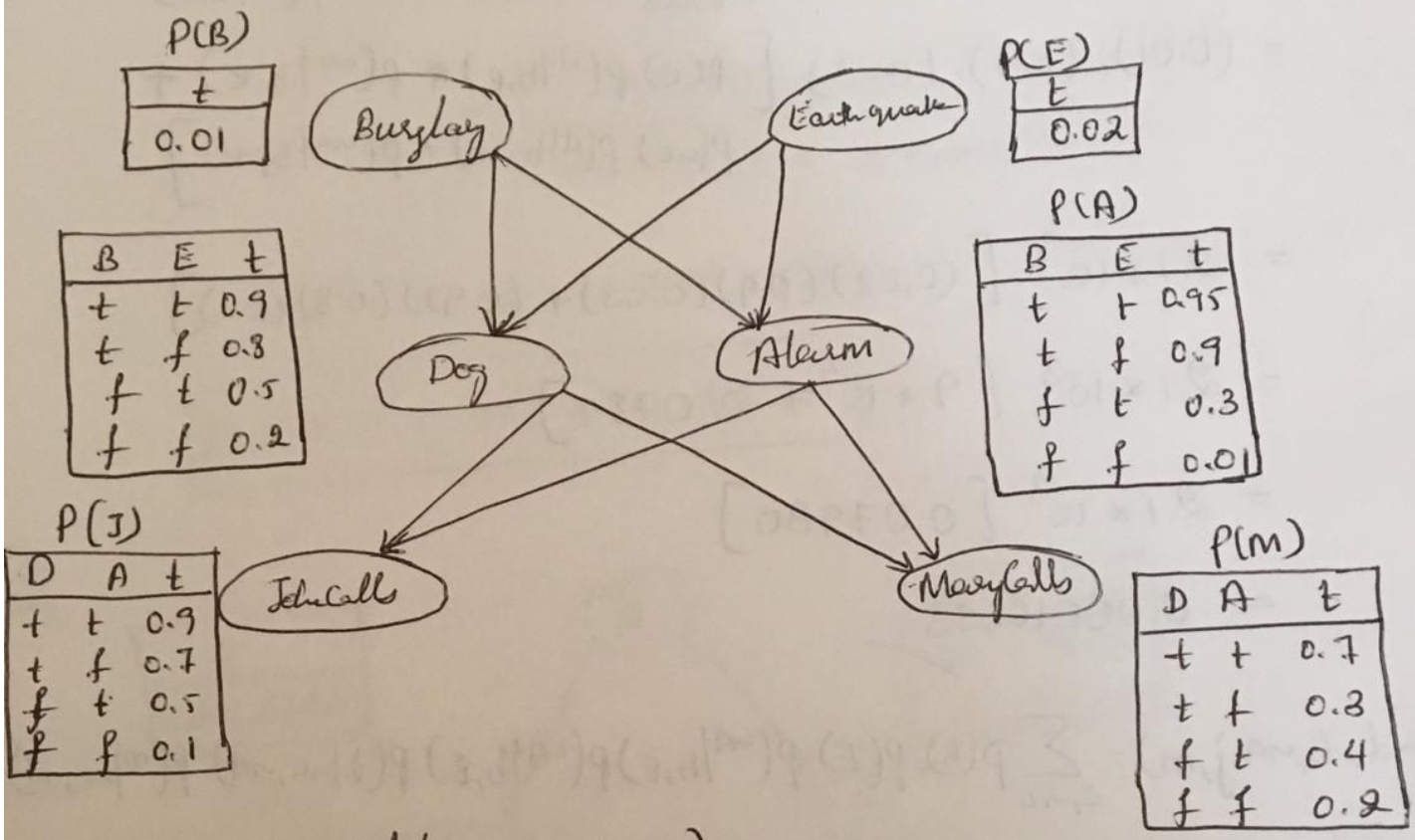
HomeWork-2

Sushma Kudlur Niranjappa
9079813292

(1)



Compute $p(d \mid b, \neg a, j, m)$.

$$\Rightarrow p(d \mid b, \neg a, j, m) = \frac{p(d, b, \neg a, j, m)}{p(b, \neg a, j, m)}$$

$$= \frac{p(d, b, \neg a, j, m)}{p(d, b, \neg a, j, m) + p(\neg d, b, \neg a, j, m)}$$

$$P(d, b, \neg a, j, m) = \sum_{e, \neg e} P(b)\, P(E)\, P(d/b, E)\, P(\neg a/b, E)\, P(j/\neg a, d)\, P(m/\neg a, d)$$

$$= P(b) \cdot P(j/\neg a, d)\, P(m/\neg a, d) \sum_{e, \neg e} P(E) \cdot P(d/b, E) \cdot P(\neg a/b, E)$$

$$= (0.01) \cdot (0.7) \cdot (0.3) \left[P(e)\, P(d/b, e) * P(\neg a/b, e) + P(\neg e)\, P(d/b, \neg e) * P(\neg a/b, \neg e \right]$$

$$= 2.1 * 10^{-3} \left[(0.02)(0.9)(0.05) + (0.98)(0.8)(0.1) \right]$$

$$= 2.1 * 10^{-3} \left[9 * 10^{-4} + 0.0784 \right]$$

$$= 2.1 * 10^{-3} \left[0.07930 \right]$$

$$= 0.00016653$$

$$P(\neg d, b, \neg a, j, m) = \sum_{e, \neg e} P(b)\, P(E)\, P(\neg d/b, E)\, P(\neg a/b, E)\, P(j/\neg a, \neg d)\, P(m/\neg a, \neg d)$$

$$= P(b)\, P(j/\neg a, \neg d)\, P(m/\neg a, \neg d) \sum_{e, \neg e} P(E)\, P(\neg d/b, E)\, P(\neg a/b, E)$$

$$= P(b)\, P(j/\neg a, \neg d)\, P(m/\neg a, \neg d) \left[P(e)\, P(\neg d/b, e)\, P(\neg a/b, e) + P(\neg e)\, P(\neg d/b, \neg e)\, P(\neg a/b, \neg e) \right]$$

$$= (0.01)(0.1)(0.2) \left[(0.02) \cdot (0.1) \cdot (0.05) + (0.98) \cdot (0.2) \cdot (0.1) \right]$$

$$= 2 * 10^{-4} \left[1 * 10^{-4} + 0.0196 \right]$$

$$= 2 * 10^{-4} \left[0.0196 \right]$$

$$= 0.0394 * 10^{-4} \approx 3.94 * 10^{-6}$$

Now,

$$P(d,b,\neg a,j,m) = \frac{P(d,b,\neg a,j,m)}{P(d,b,\neg a,j,m) + P(\neg d,b,\neg a,j,m)}$$

$$= \frac{0.00016653}{0.00016653 + 3.94*10^{-6}}$$

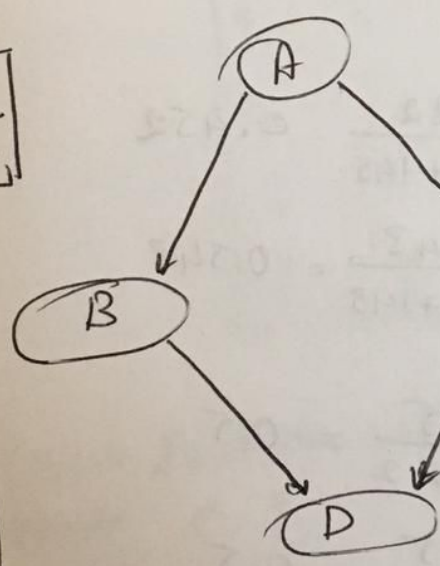
$$= 0.9768874$$

$$\boxed{P(d,b,\neg a,j,m) = 0.977}$$

(2)

(a)

A
<270, 230>



A	ψA
t	<138, 132>
f	<137, 93>

A	B|A
t	<122, 148>
f	<115, 115>

B	C	D|B,C
t	t	<66, 67>
t	f	<66, 33>
f	t	<64, 78>
f	f	<62, 59>

Computing Initial probability:

A <270, 230>

$$P(t) = \frac{270}{270+230} = 0.54 \qquad P(f) = \frac{230}{270+230} = 0.46$$

C

If A = t,

$$P_c(t) = \frac{138}{138+132} = 0.511$$

$$P_c(f) = \frac{132}{138+132} = 0.489$$

If A = f,

$$P_c(t) = \frac{137}{137+93} = 0.595$$

$$P_c(f) = \frac{93}{137+93} = 0.404$$

B If A = t,

$$P_b(t) = \frac{122}{122+148} = 0.452$$

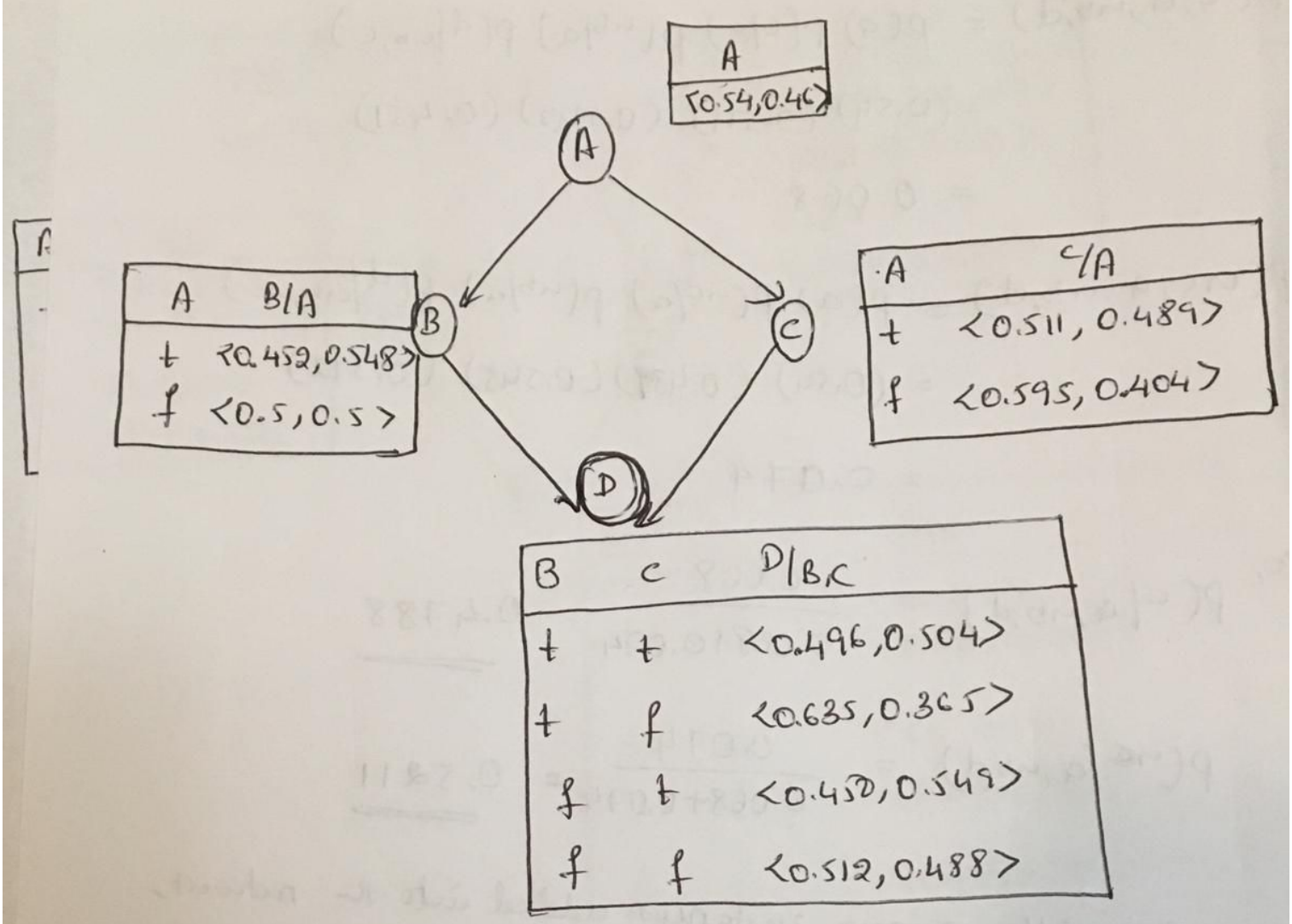
$$P_b(f) = \frac{148}{122+148} = 0.548$$

If A = f, $P_b(t) = \frac{115}{115+2} = 0.5$

$$P_b(f) = \frac{115}{115+2} = 0.5$$

D

B	C	D/B,C	Counts	probability
t	t	<66,67>	$\left\langle \frac{66}{66+67}, \frac{67}{66+67} \right\rangle$	<0.496, 0.504>
t	f	<66,38>	$\left\langle \frac{66}{66+38}, \frac{38}{66+38} \right\rangle$	<0.635, 0.365>
f	t	<64,78>	$\left\langle \frac{64}{64+78}, \frac{78}{64+78} \right\rangle$	<0.450, 0.549>
f	f	<62,59>	$\left\langle \frac{62}{62+59}, \frac{59}{62+59} \right\rangle$	<0.512, 0.488>



(2)

(b) Given: 10 new instance with A= true, B=false, d=true
estimate C= ?

Let's ~~predi~~ calculate $P(c/a, \neg b, d)$ with existing conditional probability table

$$P(c/a,\neg b,d) = \frac{P(c,a,\neg b,d)}{P(a,\neg b,d)}$$

$$= \frac{P(c,a,\neg b,d)}{P(c,a,\neg b,d) + P(\neg c,a,\neg b,d)}$$

$$P(c, a, \neg b, d) = P(a)\, P(c/a)\, P(\neg b/a)\, P(d/\neg b, c)$$

$$= (0.54)\,(0.511)\,(0.548)\,(0.451)$$

$$= 0.068$$

$$P(\neg c, a, \neg b, d) = P(a)\, P(\neg c/a)\, P(\neg b/a)\, P(d/\neg b, \neg c)$$

$$= (0.54)\,(0.489)\,(0.548)\,(0.512)$$

$$= 0.074$$

So, $P(c/a, \neg b, d) = \dfrac{0.068}{0.068+0.074} = 0.4788$

$$P(\neg c/a, \neg b, d) = \dfrac{0.074}{0.068+0.074} = 0.5211$$

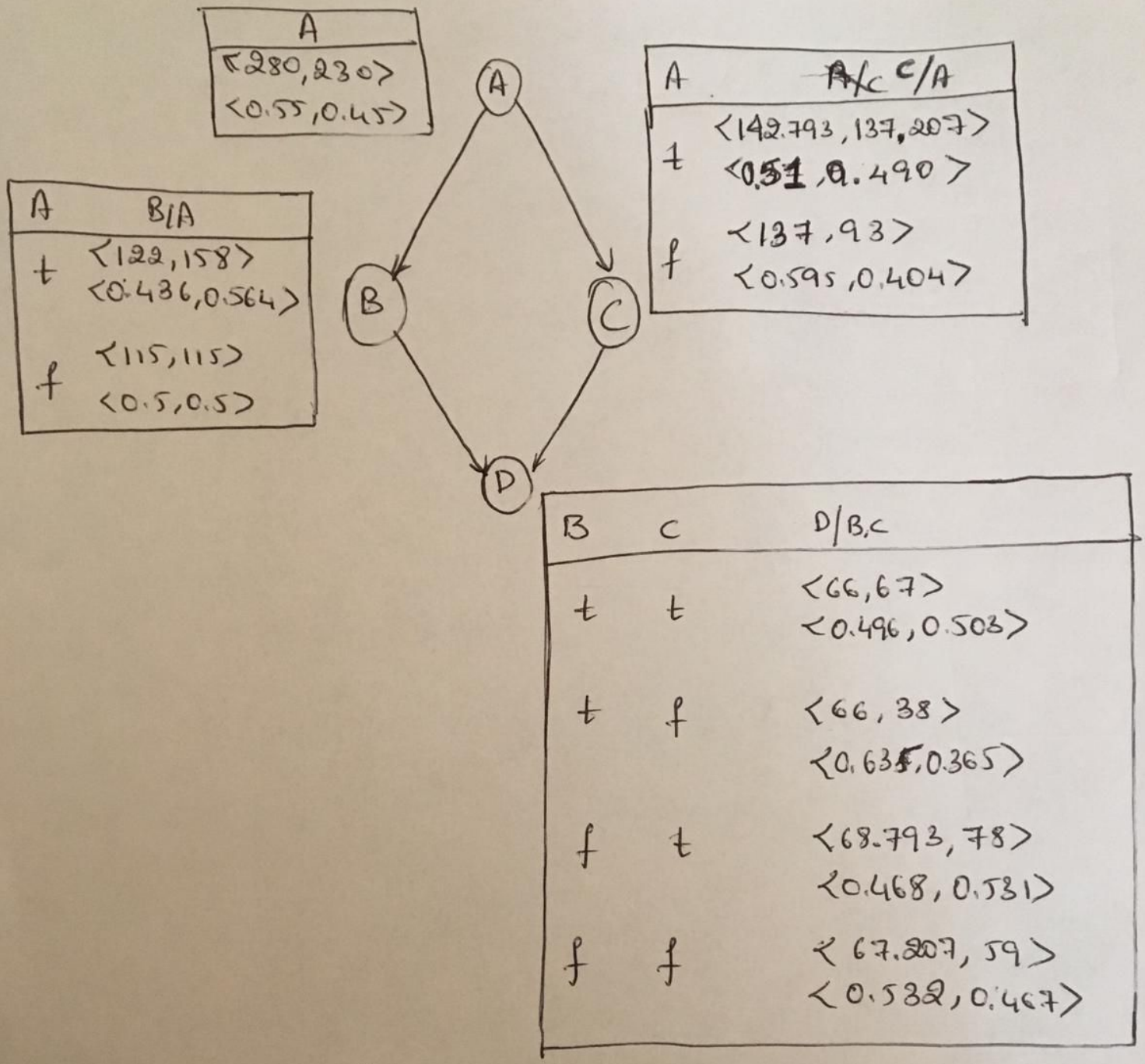
So, with 10 new instances added into the network,

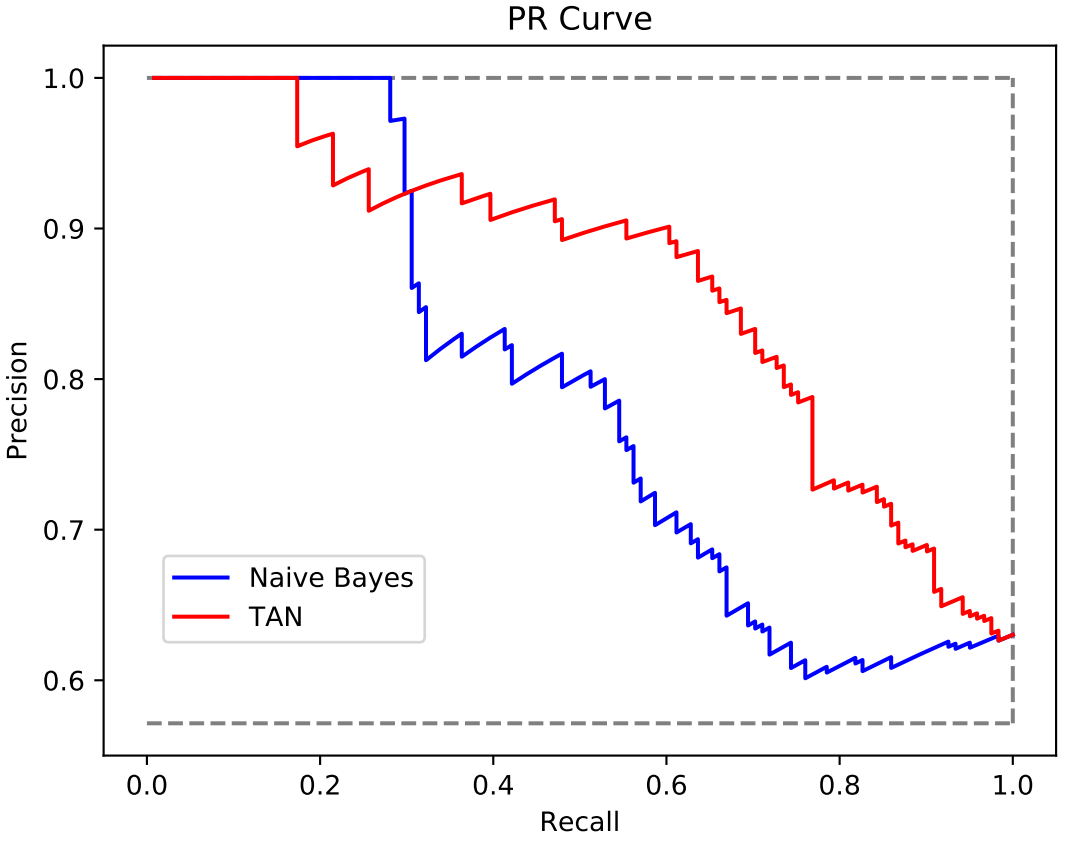
$P_c(t) = 0.4788 * 10 = 4.788$

$P_c(f) = 0.5211 * 10 = 5.211$

are the new sharel added to the exi counts.

Hence the new conditional probability table after 1-iteration of E-M will be as shown below:


A
<280,230>
<0.55,0.45>
A
B
C
D
A
B|A
t
<122,158>
<0.436,0.564>
f
<115,115>
<0.5,0.5>
A
A/C C/A
t
<142.793, 137, 207>
<0.51, 0.490>
f
<137, 93>
<0.595, 0.404>
B
C
D/B,C
t t
<66,67>
<0.496, 0.503>
t f
<66, 38>
<0.635, 0.365>
f t
<68.793, 78>
<0.468, 0.531>
f f
<67.207, 59>
<0.532, 0.467>


PR Curve
Precision
Recall
1.0
0.9
0.8
0.7
0.6
0.0
0.2
0.4
0.6
0.8
1.0
Naive Bayes
TAN

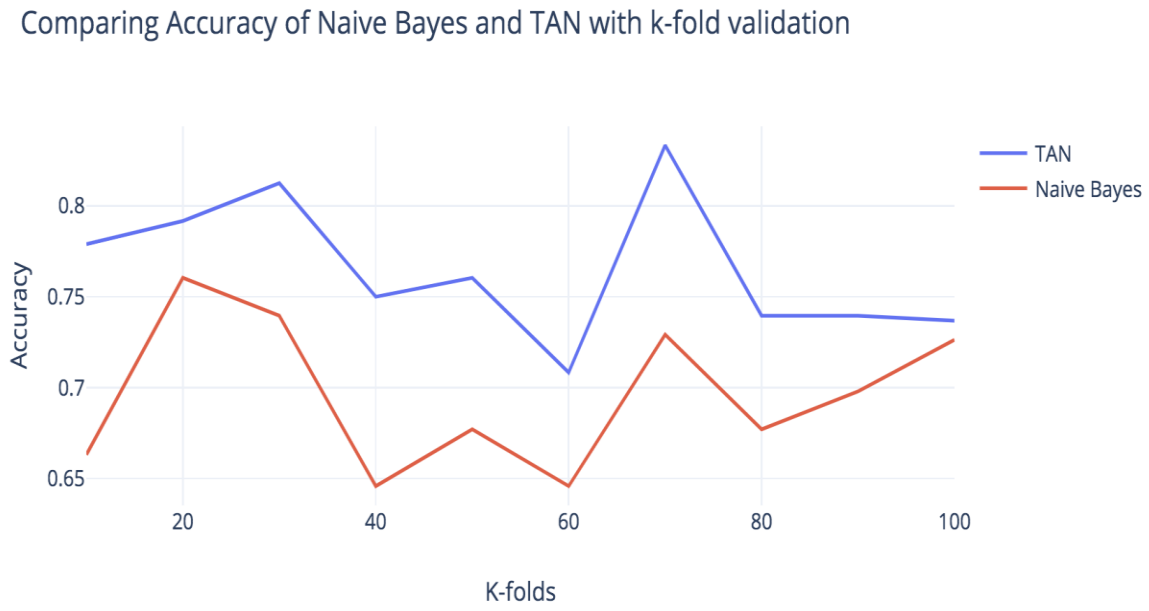
Part: 2 Precision-Recall curve:

We know that, a model with higher precision relates to low false positive rate and a model with higher recall relates to low false negative rate. In an ideal case it is expected that a model has good recall as false negatives are more dangerous that false positive (Ex: If an health check report is false negative, the patient may go untreated, which is about question of one's life). So, a model with higher recall is always the best (not that precision being worst is fine, a decent precision would still be expected).

After observing the P-R curve for both Naive-bayes & TAN, we can clearly see that TAN offered higher recall than Naive-bayes. Precision of TAN is comparitively better than Naive-bayes while we can see recall by TAN is substantially higher than Naive-bays. Hence TAN model is better and has more predictive power.

Thus a very high recall and better precision offered by TAN (compared with Naive bayes) makes it a more powerful predictive model.

Part-3



Calculated value:

Sample Mean : -0.06888157894736842

SD : 0.032494820478905485

t statistic : -6.703304560300331

P-value : 8.820443574104684e-05

Given Threshold: 0.05

In a statistical hypothesis test, a p-value would help us **to determine the significance of the result**. P-value is used to show if there is any significant difference between the systems. This offers a means to reject points to provide smallest level of significance at which the null hypothesis would be rejected. (NULL hypothesis: the hypothesis that there is no significant difference between specified populations). A small p-value (less than 0.05) suggests that null hypothesis is to be rejected while a large p-value (greater than 0.05) denotes that null hypothesis is to be accepted due to lack of counter proposition against it.

Here we see that the p-value is much smaller than 0.05. Hence **P-value helps us to show that there is a significant difference between the system.**