Home Work - 4

Prayag Nikul Purani (017416737)

Question 1

a) P(vomiting = true)

$= \frac{6}{10} = 0.6$

b) P(headache = false)

$= \frac{3}{10} = 0.3$

c) P(vomiting = false | headache = true)

× P(vomiting = false)

$= \frac{1}{4} \times \frac{4}{10} = 0.1$

d) P(Vomiting = false | headache = true)

$\frac{1}{7} = 0.1429$

e) P(Meningitis | fever = true, vomiting = f)

→ P(Meningitis = true | fever = true, vomiting = f)

$= \frac{1}{4} = \underline{0.25}$

→ P(Meningitis = false | fever = true, vomiting = f)

$= \frac{3}{4} = \underline{0.75}$

$\Rightarrow \underline{< 0.25, 0.75 >}$

Question 2

a) finding prior of target feature

P(status = ok) = 4/13 = 0.3077

P(status = settler) = 5/13 = 0.3846

P(status = solids) = 4/13 = 0.3077

$P(Ss-In \mid Ok) = (\mu = 189 \quad \sigma = 45.42)$ (2)

$P(Sed-In \mid Ok) = (\mu = 3.125 \quad \sigma = 0.25)$

$P(bond-In \mid Ok) = (\mu = 1860.5 \quad \sigma = 371.4)$

$P(Ss-Out \mid Ok) = (\mu = 18 \quad \sigma = 6.06)$

$P(Sed-Out \mid Ok) = (\mu = 0.054 \quad \sigma = 0.10)$

$P(bond-Out \mid Ok) = (\mu = 2036 \quad \sigma = 532.19)$

$P(Ss-In \mid settled) = (\mu = 200.8 \quad \sigma = 55.13)$

$P(Sed-In \mid settled) = (\mu = 4.4 \quad \sigma = 1.78)$

$P(bond-In \mid settled) = (\mu = 1251.2 \quad \sigma = 116.24)$

$P(Ss-Out \mid settled) = (\mu = 98 \quad \sigma = 23.38)$

$P(Sed-Out \mid settled) = (\mu = 1.018 \quad \sigma = 1.53)$

$P(cond-Out \mid settled) = (\mu = 1372 \quad \sigma = 142.58)$

$P(Ss-In \mid solids) = (\mu = 1301 \quad \sigma = 485.44)$

$P(Sed-In \mid solids) = (\mu = 32.5 \quad \sigma = 11.96)$

$P(bond-In \mid solids) = (\mu = 1621 \quad \sigma = 453.04)$

$P(Ss-Out \mid solids) = (\mu = 49.1 \quad \sigma = 37.76)$

$P(Sed-Out \mid solids) = (\mu = 1293 \quad \sigma = 430.95)$

$P(cond-Out \mid solids) = (\mu = 832.85 \quad \sigma = 958.81)$

b) $P(ok) = 0.3077$

$P(SS\text{-}in \mid ok) = 222 \longrightarrow 0.0068$

$P(Sed\text{-}in \mid ok) = 4.5 \longrightarrow 4.3 \times 10^{-7}$

$P(cons\text{-}in \mid ok) = 1518 \longrightarrow 0.0007$

$P(Ss\text{-}out \mid ok) = 74 \longrightarrow 1.76 \times 10^{-20}$

$P(Sed\text{-}out \mid ok) = 0.25 \longrightarrow 0.54$

$P(cons\text{-}out \mid ok) = 1642 \longrightarrow 0.0006$

$\therefore P(v[k] \mid ok) \times P(ok)$

$= 3.41577 \times 10^{-36}$

$P(settler) = 0.3846$

$P(SS\text{-}in \mid settler) = 222 \longrightarrow 0.0067$

$P(Sed\text{-}in \mid settler) = 4.5 \longrightarrow 0.2235$

$P(cons\text{-}in \mid settler) = 1518 \longrightarrow 0.0002$

$P(Ss\text{-}out \mid settler) = 74 \longrightarrow 0.0101$

$P(Sed\text{-}out \mid settler) = 0.25 \longrightarrow 0.2303$

$P(cons\text{-}out \mid settler) = 1642 \longrightarrow 0.0005$

$\therefore (P(v[k] \mid settler)) \times P(settler)$

$= 1.53837 \times 10^{-13}$

$P(\text{solids}) = 0.3077$

$P(\text{Ss-in} \mid \text{solids}) = 222 \rightarrow 6.9496 \times 10^{-5}$

$P(\text{Sed-in} \mid \text{solids}) = 4.5 \rightarrow 0.0022$

$P(\text{cond-in} \mid \text{solids}) = 1.518 \rightarrow 0.0009$

$P(\text{Ss-out} \mid \text{solids}) = 74 \rightarrow 0.0085$

$P(\text{Sed-out} \mid \text{solids}) = 0.25 \rightarrow 1.0291 \times 10^{-5}$

$P(\text{cond-out} \mid \text{solids}) = 1642 \rightarrow 0.0003$

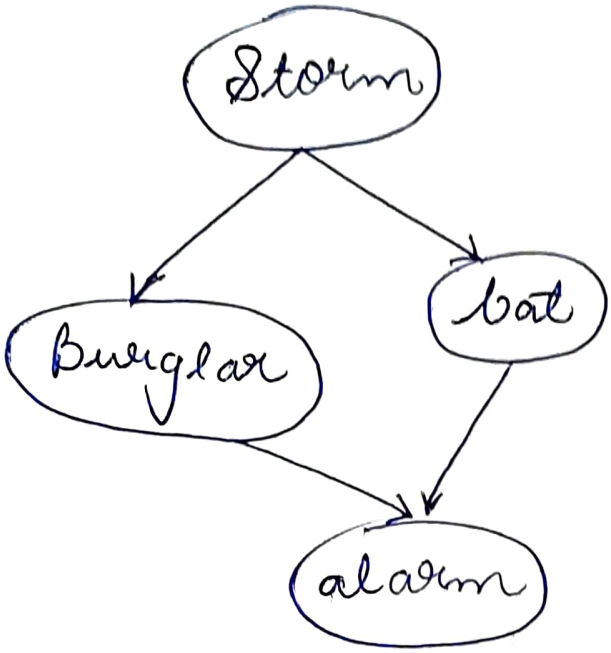
$f_{T1}(P(q[k]) \mid \text{solid}) \times P(\text{solids})$

$= 1.0068 \times 10^{-21}$

Recall that because we are using the heights of the pdf rather than calculating the actual probabilities for each feature taking a value, the score of each lever is a relative ranking is Status = settler. This indicates that there was a problem with the plant settler equipment on the day of query.

Question 3

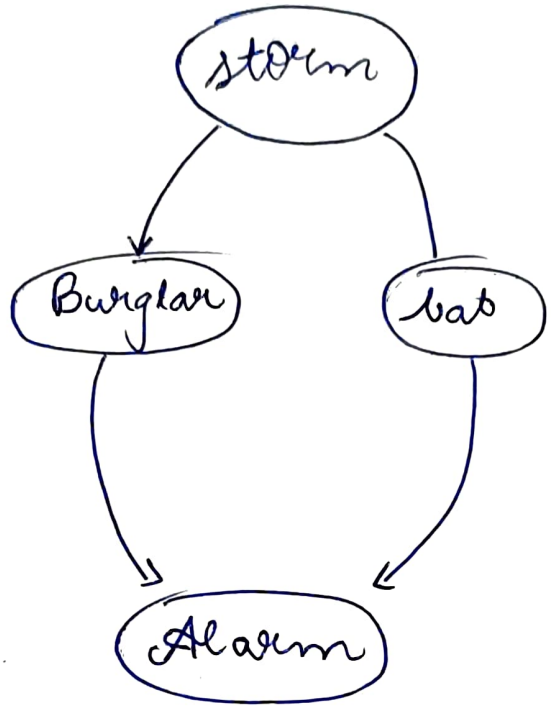
a)



The figure below illustrate a bayesian network that encodes the details causal relationship. Storms directly affect the behavior of burglar and its if this is reflected by link from the storm node to the burglar & its nodes. The behavior of burglars & its both affect whether the alarm goes off & hence there are link from each these nodes to the alarm node.

b) Conditional probability (CPT)

$$P(S=T) = 4/13 = 0.307$$



$P(B=T|S)$

S	P
T	1/4 = 0.25
F	3/9 = 0.33

$P(cab=T|S)$

S	
T	3/4 = 0.75
F	2/9 = 0.222

B	C	$P(A=T\|B,C)$
T	T	1/1 = 1
T	F	2/3 = 0.664
F	T	1/4 = 0.25
F	F	1/5 = 0.2

c) from the CPT table we can see that when Burglar = true, bab = true & storm = false then storm = true - this the story will go on for the above situation

d) $P(a|s) = \frac{P(a,s)}{P(s)}$

$$= \sum_{ij} \frac{P(a, B_i, C_j, s)}{P(s)}$$

$$\sum P(a,B,C,s) = \sum P(a|B,C) \times P(B|s) \times P(C|s) \times P(s)$$

$$(1 \times 0.25 \times 0.75 \times 0.3077)$$
$$+$$
$$(0.666 \times 0.25 \times 0.25 \times 0.3077)$$
$$+$$
$$(0.25 \times 0.75 \times 0.75 \times 0.3077)$$
$$+$$
$$(0.2 \times 0.75 \times 0.25 \times 0.3077)$$

$$= \frac{0.125324}{0.3077} = \underline{0.4073}$$

This implies $P(\text{alarm} = \text{false}) = \underline{0.5927}$

Question 4

5

a) sorting age

18, 19, 21, 43, 41, 49, 51, 55, 57

3 bins of 9 records

$9/3 = 3$ bins

bin young → ID 9, ID 6, ID 3

bin middle → ID 1, ID 4, ID 7

bin mature → ID 8, ID 5, ID 2.

$$\text{threshold} = \frac{21+43}{2} = 32$$

$$\text{threshold} = \frac{49+51}{2} = 50.$$

b) The features from which no useful information can be extracted or those cant be used to predict target should be excluded eg is name should be excluded as it unique & doesn't hold any relavance but 'occupation'

feature should also be excluded as it is unique in this case & can't help for prediction problem.

c) Naive Bayes model

P(phone) = 0.56

P(gender = female | phone) = 3/5

P(gender = male | phone) = 2/5

P(age = young | phone) = 1/5

P(age = middle-age | phone) = 2/5

P(age = mature | phone) = 2/5

P(policy = A | phone) = 1/5

P(policy = B | phone) = 1/5

P(policy = C | phone) = 3/5

P(email) = 0.44

P(gender = female | phone) = 1/4

P(gender = male | phone) = 3/4

P(age = young | phone) = 2/4

P(age = middle-age | phone) = 1/4

P(age = mature | phone) = 2/4

P(policy = A | phone) = 2/4

P(policy = B | phone) = 1/4

P(policy = C | phone) = 1/4

d) gender = female, age = 30, policy = A

(i) (0.56 × 0.6 × 0.2 × 0.2)

= 0.0134

(ii) (0.44 × 0.25 × 0.5 × 0.5)

= 0.0275

So, it is quite evidant that value obtain from above that the prediction model predict the value of target feature as channel = Email

Question 5

a)

~~Education~~ Entertainment

fun = $\frac{415}{700}$ = 0.593

is = $\frac{695}{700}$ = 0.99

learning = $\frac{35}{700}$ = 0.05

machine = $\frac{70}{700}$ = 0.10

Education

fun = $\frac{200}{300}$ = 0.667

is = $\frac{295}{300}$ = 0.983

learning = $\frac{120}{300}$ = 0.40

machine = $\frac{105}{300}$ = 0.35

P(entertainment | q) =

(0.7 × 0.593 × 0.99 × 0.025 × 0.10)

= 0.00205

P(Education | q) =

(0.3 × 0.667 × 0.983 × 0.4 × 0.35)

= 0.00275

P(Education) > P(entertainment)

b) q = "christmas family fun"

where observing words of query we found that word "christmas didn't appear in either of the dataset, so conditional probability for this will be 0.

P(christmas | entertainment) = 0

P(christmas | education) = 0

c) entertainment

raw probability

$P(c|e) = 0$

$P(\text{family}|e) = 0.5714$

$P(\text{fun}|e) = 0.5929$

Smoothing parameter

$k = 10$, vocabulary = 6, count = 700

christmas = 0, family = 400,

fun = 415

Smoothing probability

$P(\text{christmas}|\text{entertainment}) = \frac{0+10}{700+(10\times6)} = 0.0132$

$P(\text{family}|\text{entertainment}) = \frac{400+10}{700+(10\times6)} = 0.5395$

$P(\text{fun}|\text{entertainment}) = \frac{415+10}{700+(10\times6)} = 0.5592$

education

Raw probabilities

P(c | edu) = 0

P(family | edu) = 0.5714

P(fun | edu) = 0.5929

Smoothing parameter

k = 10 education = 300, christmas = 0

family = 10, fun = 200,

word = 6.

Smoothed probabilities

P(christmas | entertainment)

$$= \frac{0 + 10}{300 + (10 \times 6)} = 0.0278$$

P(family | entertainment)

$$= \frac{10 + 10}{300 + (10 \times 6)} = 0.0556$$

P(fun | entertainment)

$$= \frac{200 + 10}{300 + (10 \times 6)} = 0.5833$$