Homework-8 (CS5805 Machine Learning)

Jyothi Sevakula.

Q1) (a) $P(A=0|+)=?$

$$P(A=0|+) = \frac{P(A=0,+)}{P(+)}$$

$$= \frac{2/10}{5/10} = 2/5 = 0.4$$

$$\boxed{P(A=0|+)=0.4}$$

(b) $P(A=0|-)=?$

$$P(A=0|-) = \frac{P(A=0,-)}{P(-)}$$

$$= \frac{3/10}{5/10} = 0.6$$

$$\boxed{P(A=0|-) = 0.6}$$

(c) $P(B=1|+)=?$

$$P(B=1|+) = \frac{P(B=1,+)}{P(+)}$$

$$= \frac{1/10}{5/10} = 0.2$$

$$\boxed{\cancel{P(B|+)}\ P(B=1|+) = 0.2}$$

(d) $P(B=1|-)=?$

$$P(B=1|-) = \frac{P(B=1,-)}{P(-)}$$

$$= \frac{2/10}{5/10} = 0.4$$

$$\boxed{P(B=1|-)=0.4}$$

(e) $P(C=0|+)$

$$P(C=0|+) = \frac{P(C=0,+)}{P(+)}$$

$$= \frac{1/10}{5/10} = 0.2$$

$$\boxed{P(C=0|+) = 0.2}$$

(f) $P(C=0|-)$

$$P(C=0|-) = \frac{P(C=0,-)}{P(-)}$$

$$= \frac{0/10}{5/10} = 0$$

$$\boxed{P(C=0|-)=0}$$

Q2) Given test sample $(A=0, B=1, C=0)$

$$\cancel{P(yes|x) = \frac{P(x|yes)\cdot P(yes)}{P(x)}}$$

$$P(+|X) = \frac{P(X|+)\cdot P(+)}{P(X)}$$

$$P(X|+) = P(X_1, X_2, X_3|+) = P(X_1|+) * P(X_2|+) * P(X_3|+)$$

we can say, $\xrightarrow{}$ proportionality because denominator $P(X)$ is common for both cases $P(+|X)$ and $P(-|X)$

$P(+|X) \propto P(x_1|+) \times P(x_2|+) \times P(x_3|+) \times P(+)$

$\Rightarrow \frac{2}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{5}{10}$

$\Rightarrow \frac{2}{250}$

$P(-|X) \propto P(x_1|-) \times P(x_2|-) \times P(x_3|-) \times P(-)$

$\Rightarrow \frac{3}{5} \times \frac{2}{5} \times 0 \times \frac{5}{10}$

$\Rightarrow 0$

$P(+|X) > P(-|X) \Rightarrow$ where $X = (A=0, B=1, C=0)$ can be <u>classified as '+'</u>

Q3) Laplace smoothing, $P(A_i|c) = \frac{N_{ic} + \alpha}{N_c + p \times \alpha}$ ($\alpha = 2$ given, $p=2$ for all A, B, C i.e, no of classes $\Rightarrow$ 0 or 1)

a) $P(A=0|+) = \frac{count(A=0,+) + 2}{count(+) + 2\times2}$

$= \frac{2+2}{5+4} = \frac{4}{9}$

$\boxed{P(A=0|+) = 4/9}$

b) $P(A=0|-) = \frac{count(A=0,-)+2}{count(-) + 4} = \frac{3+2}{9} = \frac{5}{9}$

$\boxed{P(A=0|-) = 5/9}$

c) $P(B=1|+) = \frac{count(B=1,+)+2}{count(+)+(2\times2)} = \frac{1+2}{5+4} = \frac{3}{9}$

$\boxed{P(B=1|+) = 1/3}$

Scanned with CamScanner

d) $P(B=1|-) = \frac{count(B=1,-)+2}{count(-)+4} = \frac{2+2}{9} = \frac{4}{9}$

$\boxed{P(B=1|-) = \frac{4}{9}}$

e) $P(C=0|+) = \frac{count(C=0,+)+2}{count(+)+4} = \frac{1+2}{5+4} = \frac{3}{9}$

$\boxed{P(C=0|+) = 1/3}$

f) $P(C=0|-) = \frac{count(C=0,-)+2}{count(-)+4} = \frac{0+2}{5+4} = 2/9$

$\boxed{P(C=0|-) = 2/9}$

Q4) given test sample $A=0, B=1, C=0$

considered probabilities after applying laplace smoothing

$P(+|A=0,B=1,C=0) \propto P(A=0|+) \cdot P(B=1|+) \cdot P(C=0|+) \cdot P(+)$

$\Rightarrow \frac{4}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{1}{2}$

$\Rightarrow \frac{36}{1458}$

$P(-|A=0,B=1,C=0) \propto P(A=0|-) \cdot P(B=1|-) \cdot P(C=0|-) \cdot P(-)$

$\Rightarrow \frac{5}{9} \cdot \frac{4}{9} \cdot \frac{2}{9} \cdot \frac{1}{2}$

$\Rightarrow \frac{40}{1458}$

$P(-|A=0,B=1,C=0) > P(+|A=0,B=1,C=0)$

$\therefore$ Test record $A=0, B=1, C=0$ can be classified as '−'

Q5)) The label prediction for test sample, $A=0, B=1, C=0$

without laplace smoothing (Q2) $\Rightarrow$ classified as '+'

with laplace smoothing (Q4) $\Rightarrow$ classified as '−'

The results are not same in q2 and q4.

I prefer laplace smoothing method, as it prevents zero probabilities. without smoothing, if any conditional probability is 0 (due to feature value being absent for a class ; in q2 $P(C=0|-)$), it causes the entire posterior probability for that class to become 0 (In q2, because of $P(C=0|-) \Rightarrow P(-|A=0, B=1, C=0)$ is 0). This can lead to incorrect predictions even if other feautures strongly support that class. Laplace smoothing prevent this by adding small constant (α) to all the feature counts in the formula for conditional probabilities.

Q6)

a) $P(A=1|+)$

$$= \frac{P(A=1,+)}{P(+)} = \frac{3/10}{5/10}$$

$$\boxed{P(A=1,+) = 3/5}$$

b) $P(A=1|-)$

$$= \frac{P(A=1,-)}{P(-)} = \frac{2/10}{5/10}$$

$$\boxed{P(A=1|-) = 2/5}$$

c) $P(B=1|+)$

$$= \frac{P(B=1,+)}{P(+)} = \frac{2/10}{5/10}$$

$$\boxed{P(B=1|+) = 2/5}$$

d) $P(B=1|-)$

$$= \frac{P(B=1,-)}{P(-)} = \frac{2/10}{5/10}$$

$$\boxed{P(B=1|-) = 2/5}$$

e) $P(C=1|+)$

$$= \frac{P(C=1,+)}{P(+)} = \frac{4/10}{5/10}$$

$$\boxed{P(C=1|+) = 4/5}$$

f) $P(C=1|-)$

$$= \frac{P(C=1,-)}{P(-)} = \frac{1/10}{5/10}$$

$$\boxed{P(C=1|-) = 1/5}$$

Q7) given test sample (A=1, B=1, C=1)

if probability of $P(+ \mid A=1, B=1, C=1) > P(- \mid A=1, B=1, C=1)$

we can classify as '+'.

$$P(+ \mid A=1, B=1, C=1) \propto P(A=1 \mid +) \cdot P(B=1 \mid +) \cdot P(C=1 \mid +) \cdot P(+)$$

directly proportional because we are neglecting denominator $P(x)$ for both cases $P(+ \mid x)$ and $P(- \mid x)$ where $x = (A=1, B=1, C=1)$

$$\Rightarrow \frac{3}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{5}{10}$$

$$\Rightarrow \frac{3 \times 2 \times 4}{250} = \frac{24}{250}$$

$$P(- \mid A=1, B=1, C=1) \propto P(A=1 \mid -) \cdot P(B=1 \mid -) \cdot P(C=1 \mid -) \cdot P(-)$$

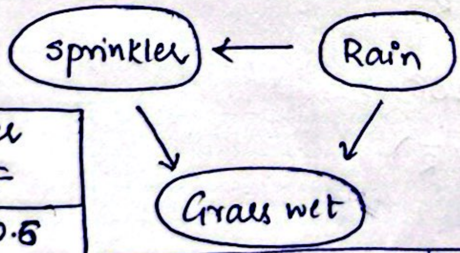
$$\Rightarrow \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{\cancel{5}}{\cancel{10}\,2}$$

$$\Rightarrow \frac{4}{250}.$$

$$\therefore P(+ \mid A=1, B=1, C=1) > P(- \mid A=1, B=1, C=1)$$

From above, we can classify test sample (A=1, B=1, C=1) as "+"

Q8)



	Rain	
	T	F
	0.2	0.8

	sprinkler	
Rain	T	F
F	0.4	0.6
T	0.01	0.99

grass wet

sprinkler	rain	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Given that grass is wet, we need to find out it rained or not ?

$$P(\text{Rain} \mid \text{Grass wet}) = \frac{P(\text{Grasswet} \mid \text{Rain}) \times P(\text{rain})}{P(\text{grass wet})}$$

$$P(\text{rain, sprinkler, grass wet}) = P(\text{grasswet} \mid \text{sprinkler, rain}) \times P(\text{sprinkler} \mid \text{rain}) \times P(\text{rain})$$

$$= 0.99 \times 0.01 \times 0.2$$

$$\boxed{P(R,S,G) = 0.00198}$$

so we have 3 random variables R, S, G hence $2^3 = 8$ combination

$P(R,S,G)$	S,G	$\bar{S},G$	$S,\bar{G}$	$\bar{S},\bar{G}$
R	0.00198	0.1584	0.00002	0.0396
$\bar{R}$	0.288	0	0.032	0.48

$$P(R,\bar{S},G) = P(G \mid \bar{S},R) \times P(\bar{S} \mid R) \times P(R)$$

$$= 0.8 \times 0.99 \times 0.2 = 0.1584$$

$$P(R,S,\bar{G}) = P(\bar{G} \mid S,R) \times P(S \mid R) \times P(R)$$

$$= 0.01 \times 0.01 \times 0.2 = 0.00002.$$

$$P(R,\bar{S},\bar{G}) = P(\bar{G} \mid R,\bar{S}) \times P(\bar{S} \mid R) \times P(R)$$

$$= 0.2 \times 0.99 \times 0.2 = 0.0396$$

$$P(\bar{R},S,G) = P(G \mid \bar{R},S) \times P(S \mid \bar{R}) \times P(\bar{R})$$

$$= 0.9 \times 0.4 \times 0.8 = 0.288$$

$$P(\bar{R},\bar{S},G) = P(G \mid \bar{R},\bar{S}) \times P(\bar{S} \mid \bar{R}) \times P(\bar{R})$$

$$= 0 \times P(\bar{S} \mid \bar{R}) \times P(\bar{R}) = 0$$

$$P(\bar{R},S,\bar{G}) = P(\bar{G} \mid \bar{R},S) \times P(S \mid \bar{R}) \times P(\bar{R})$$

$$= 0.1 \times 0.4 \times 0.8 = 0.032$$

$$P(\bar{R},\bar{S},\bar{G}) = P(\bar{G} \mid \bar{R},\bar{S}) \times P(\bar{S} \mid \bar{R}) \times P(\bar{R})$$

$$= 1 \times 0.6 \times 0.8 = 0.48$$

	G	
R	0.16038	$(P(R,S,G)+P(R,\bar{S},G))$
$\bar{R}$	0.288	$(P(\bar{R},S,G)+P(\bar{R},\bar{S},G))$

$$P(G) = 0.16038 + 0.288 = 0.44838$$

$$P(R|G) = \frac{P(R,G)}{P(G)} = \frac{0.16038}{0.44838} = 0.357 \approx 36\%$$

$$P(\bar{R}|G) = \frac{P(\bar{R},G)}{P(G)} = \frac{0.288}{0.44838} = 0.642 \approx 64\%$$

$$P(\bar{R}|G) > P(R|G)$$

given grass is wet it is more likely that "<u>it did not rain</u>"
as $P(\bar{R}|G) > P(R|G)$.