Homework No. 3.

Name : Kavya Kodishala

UNTID: 11614043

Course Number: CSCE 5380-002

1) a) Given that,

the fraction of undergraduage students who smoke is 15%

$$P(\text{Smoker} | \text{undergraduate}) = 15\% = 0.15$$

The fraction of graduate students who smoke is 25%.

$$P(\text{Smoker} | \text{graduate}) = 25\% = 0.25$$

Assuming that one-fifth of the college students are graduate

$$p(\text{graduate}) = 1/5 = 0.2$$

$$p(\text{undergraduate}) = 4/5 = 0.8$$

$$p(G \cap S) = 0.2 * 0.25 = 0.05$$

$$p(S) = (0.2 * 0.25) + (0.15 * 0.8)$$

$$= 0.05 + 0.12$$

$$= 0.17$$

$$P(\text{graduate} | \text{smoker}) = \frac{0.05}{0.17} = 0.294 \simeq 29\%$$

2) a) If two events x_1 and x_2 are independent on each other then.

$$p(x_1 \cap x_2) = p(x_1) \cdot p(x_2)$$

Considering x_2, x_1 are two events from positive outcomes.

$x_1 \backslash x_2$	1	0	
1	20/50	20/50	40/50
0	5/50	5/50	10/50
	25/50	25/50	

$$p(x_1=1 \cap x_2=1) = \frac{20}{50} = \frac{2}{5}$$

$$P(x_1=1) \cdot p(x_2=1) = \frac{40}{50} * \frac{25}{50} = \frac{2}{5}$$

$$P(x_1=1 \cap x_2=0) = \frac{20}{50} = \frac{2}{5}$$

$$p(x_1=1) \cdot p(x_2=0) = \frac{40}{50} * \frac{25}{50} = \frac{2}{5}$$

$$p(x_1=0 \cap x_2=1) = \frac{5}{50} = \frac{1}{10}$$

$$P(x_1=0) \cdot p(x_2=1) = \frac{10}{50} * \frac{25}{50} = \frac{1}{10}$$

$p(x_1=0 \cap x_2=0) = \frac{5}{10} = \frac{1}{10}$

$p(x_1=0) \cdot p(x_2=0) = \frac{10}{50} * \frac{25}{50} = \frac{1}{10}$.

From above we can say that

$p(x_1 \cap x_2) = p(x_1) \cdot p(x_2)$ any values of 0 or 1

considering x_1, x_2 two events for negatives. outcomes

$x_1 \backslash x_2$	1	0	
1	8/50	17/50	25/50
0	8/50	17/50	25/50
	16/50	34/50	

$p(x_1=1) \cdot p(x_2=1) = 8/50$

$p(x_1=1 \cap x_2=1) = \frac{25}{50} * \frac{16}{50} = \frac{8}{50}$

$p(x_1=1 \cap x_2=0) = 17/50$

$p(x_1=1) \cdot p(x_2=0) = \frac{25}{50} * \frac{34}{50} = \frac{17}{50}$

$P(x_2 = 1 \cap x_1 = 0) = 8/50$

$p(x_1 = 0) \cdot p(x_2 = 1) = \frac{16}{50} * \frac{25}{50} = \frac{8}{50}$

$P(x_1 = 0) \cdot P(x_2 = 0) = \frac{25}{50} * \frac{34}{50} = \frac{17}{50}$

$P(x_1 = 0 \cap x_2 = 0) = 17/50.$

From above for the negative outcomes

$p(x_1 \cap x_2) = p(x_1) \cdot p(x_2)$ for any 0 & 1

$\therefore$ x_1 and x_2 are independent to each other.

3) a) conditional Probability:

$$p(x|y) = p(x,y) / p(y)$$

$$p(x,y) = p(x|y) * p(y)$$

$$= P(y|x) * p(x)$$

$$p(x_1 = 1|+) = \frac{40}{50} * \frac{1}{2} = \frac{4}{10}$$

$$p(x_1 = 1|-) = \frac{25}{50} * \frac{1}{2} = \frac{1}{4}$$

$p(x_2=1|+) = \frac{25}{50} * \frac{1}{2} = \frac{1}{4}$

$p(x_2=1|-) = \frac{16}{50} * \frac{1}{2} = \frac{4}{25}$

$p(x_3=1|+) = \frac{20}{50} * \frac{1}{2} = \frac{1}{5}$

$p(x_3=1|-) = \frac{8}{50} * \frac{1}{2} = \frac{2}{25}$

3) b)

$p(x_1=1|+) \cdot p(x_2=1|+) \cdot p(x_3=1|+)\, p(+)$

$= \frac{4}{10} * \frac{1}{4} * \frac{1}{5}$

$= \frac{1}{50} = 0.02.$

$p(x_1=1|-) \cdot p(x_2=1|-) \cdot p(x_3=1|-)\, p(-)$

$= \frac{1}{4} * \frac{4}{25} * \frac{2}{25}$

$= \frac{2}{50} = \frac{1}{25} = 0.04$

when we compare both values the negative out come is more than positive outcome.

4) a) Conditional probability.

$P(\text{accident} = \text{yes} | +) = \frac{10+10+5}{40} = \frac{25}{40} = \frac{5}{8}$

$P(\text{accident} = \text{yes} | -) = \frac{5+5+10}{60} = \frac{10}{60} = \frac{1}{6}$

$P(\text{weather} = \text{good} | +) = \frac{5+10+10}{40} = \frac{25}{40} = \frac{5}{8}$

$P(\text{weather} = \text{good} | -) = \frac{30+20+5}{60} = \frac{55}{60} = \frac{11}{12}$.

$P(\text{construction} = \text{yes} | +) = \frac{10+5}{40} = \frac{15}{40} = \frac{3}{8}$.

$P(\text{construction} = \text{yes} | -) = \frac{20+0}{60} = \frac{1}{3}$.

4) b) Given feature set:-

(Accident = no, weather = bad, construction = yes)

congestion = 40% = 0.4.

no congestion = 60% = 0.6.

From naive Bayes classifier equation:

$V_{NB} = \arg\max P(V_j) \prod_i P(a_i | V_j)$

a_i = set of attributes.

V_j = positive outcomes.

Scanned with CamScanner

positive outcome.

$P(\text{accident}|\text{NO}) = \frac{5+10}{40} = \frac{15}{40}$

$P(\text{weather}|\text{bad}) = \frac{10+5}{40} = 15/40$

$P(\text{construction}|\text{yes}) = \frac{10+5}{40} = \frac{15}{40} = \frac{15}{40}$.

Negative outcome.

$P(\text{accident}|\text{NO}) = \frac{30+20}{60} = \frac{50}{60} = \frac{5}{6}$.

$P(\text{weather}|\text{bad}) = \frac{5+0}{60} = \frac{5}{60} = \frac{1}{12}$

$P(\text{construction}|\text{yes}) = \frac{20+0}{60} = \frac{1}{3}$.

$V_{NB} = 0.4 * \frac{15}{40} * \frac{15}{40} * \frac{15}{40}$.

$= 0.02109$.

$V_{NB} = [P = NO = \text{congestion}] \cdot P[\text{accident}/NO] \cdot$

$P[\text{weather}|\text{bad}] \cdot P[\text{construction}/\text{yes}]$

$= 0.6 * \frac{5}{6} * \frac{1}{12} * \frac{1}{3}$.

$= 0.0138$

Scanned with CamScanner

$= 0.0138.$

$$V_{NB}(\text{congestion}) = \frac{0.02109}{0.02109 + 0.0138}$$

$$= \frac{0.02109}{0.03489}$$

$$= 0.604.$$

$$V_{NB}(\text{No congestion}) = \frac{0.0138}{0.02109 + 0.0138}$$

$$= \frac{0.0138}{0.03489}$$

$$= 0.395$$

$\therefore V_{NB}(\text{congestion}) > V_{NB}(\text{no congestion}).$

Scanned with CamScanner

3) a)

$p(x_1=1 \mid +) = \frac{40}{50} = 0.8$

$p(x_1=1 \mid -) = \frac{25}{50} = 0.5$

$p(x_2=1 \mid +) = \frac{25}{50} = 0.5$

$p(x_2=1 \mid -) = \frac{16}{50} = 0.32$

$p(x_3=1 \mid +) = \frac{20}{50} = 0.4$

$p(x_3=1 \mid -) = \frac{8}{50} = 0.16$

3) b)

$p(+|R) = p(x_1 = (1|+)) \cdot p(x_2 = (1|+)) \cdot p(x_3 = (1|+))$

$= \cancel{0.8} * 0.5 * 0.4$

$= 0.16$

$p(-|R) = p(x_1 = (1|-)) \cdot p(x_2 = (1|-)) \cdot p(x_3 = (1|-))$

$= 0.5 * 0.32 * 0.16 = 0.256$

$= 0.026$

$P(-|R) = 0.26$

Scanned with CamScanner

Here $P(+|R) > P(-|R)$ therefore it is classified as positive.

Example for training is $R(x_1=1, x_2=0, x_3=0)$

$P(+|R) = P(x_1=(1|+)) * P(x_2=(0|+)) * P(x_3=(0|+))$

$= 0.8 * 0.5 * 0.6$

$= 0.24$

$P(-|R) = P(x_1=(1|-)) * P(x_2=(0|-)) * P(x_3=(0|-))$

$= 0.5 * 0.68 * 0.84$

$= 0.2856$

0.286

Here $P(+|R) < P(-|R)$ therefore it is classified as negative.

For training Example R is $(x_1=0, x_2=1, x_3=0)$

$P(+|R) = P(a_1=(0|+)) * P(a_2=(1|+)) *$
$P(a_3=(0|+))$

$= 0.2 * 0.5 * 0.6$

$= 0.06.$

$P(-|R) = P(a_1=(0|-)) * P(a_2=(1|-)) *$
$P(a_3=(0|-))$

$= 0.5 * 0.32 * 0.84$

$= 0.134.$

Here $P(+|R) < P(-|R)$ Therefore it is classified as negative.

For training example R is $(a_1=0, x_2=0, x_3=0)$

$P(+|R) = P(a_1=(0|+)) * P(a_2=(0|+)) * P(a_3=(0|+))$

$= 0.2 * 0.5 * 0.6$
$= 0.06$

$P(-|R) = P(a_1=(0|-)) * P(a_2=(0|-)) * P(a_3=(0|-))$
$= 0.5 * 0.68 * 0.84$

$= 0.286$

Here $P(+|R) < P(-|R)$ therefore it is classified as negative.

Misclassified cases are:

For $x_1=1, x_2=1, x_3=1$ is 20

for $x_1=1, x_2=0, x_3=0$ is 20

for $x_1=0, x_2=1, x_3=0$ is 5

for $x_1=0, x_2=0, x_3=0$ is 5

$$\text{Training Error Rate} = \frac{\text{Misclassified class}}{\text{Total cases}}$$

$$\text{Training Error Rate} = \frac{50}{100} = 0.5$$

Scanned with CamScanner