Name: Noman khan
Seat no: B20102131

Question 4.1: Solution:

Given likelihoods:

	pos	neg
I	0.09	0.16
always	0.07	0.06
like	0.29	0.06
foreign	0.04	0.15
films	0.08	0.11

Prior Probabilities (equal):

$P(pos) = P(neg) = 0.5$

$P(pos \mid \text{'I always like foreign films'}) = P(pos) \times P(\text{'I'} \mid pos) \times P(\text{'always'} \mid pos) \times P(\text{'like'} \mid pos) \times P(\text{'foreign'} \mid pos) \times P(\text{'films'} \mid pos)$

$\Rightarrow P(pos \mid \text{'I always like foreign films'}) = 0.5 \times 0.07 \times 0.29 \times 0.04 \times 0.08$

$\Rightarrow \boxed{P(pos \mid \text{'I always like foreign films'}) = 0.00003248}$

P(neg | 'I always like foreign films') = 0.5 × 0.16 × 0.06 × 0.06 × 0.15 × 0.11

=) $\boxed{\text{P(neg | 'I always like foreign films') = 0.0000297}}$

∵ P(pos | 'I always like foreign films') > P(neg | 'I always like foreign films')

∴ Naive Bayes would classify the sentence 'I always like foreign films' as positive.

Question 4.2: Solution:

Step 1: Computing Prior Probabilities:

$P(Comedy) = \frac{2}{5} = 0.4$

$P(action) = \frac{3}{5} = 0.6$

Comedy Class:

words	Count
fun	3

Couple	2
love	2
fly	1
fast	1

Total words in Comedy class: 9

Vocabulary (unique words) size:

Action class:

Words	Count
fast	2
furious	2
Shoot	4
fun	1
fly	1
love	1

Total words in action class: 11

Vocabulary: { fun, Couple, love, fly, fast, furious, Shoot }

(words in both classes /total words)

Vocabulary Size: 7

Likelihoods with Add-1 Smoothing:

$$P(\text{word} \mid \text{Class}) = \frac{\text{Count (word in Class)} + 1}{\text{total words in Class} + \text{Vocabulary Size}}$$

Comedy Class likelihoods:

$$P(\text{fast} \mid \text{Comedy}) = \frac{1+1}{9+7} = \frac{2}{16} = 0.125$$

$$P(\text{Couple} \mid \text{Comedy}) = \frac{2+1}{9+7} = \frac{3}{16} = 0.1875$$

$$P(\text{Shoot} \mid \text{Comedy}) = \frac{0+1}{9+7} = \frac{1}{16} = 0.0625$$

$$P(\text{fly} \mid \text{Comedy}) = \frac{1+1}{9+7} = \frac{2}{16} = 0.125$$

Action Class likelihoods:

$$P(\text{fast} \mid \text{action}) = \frac{2+1}{11+7} = \frac{3}{18} = 0.166$$

$$P(\text{Couple} \mid \text{action}) = \frac{0+1}{11+7} = \frac{1}{18} = 0.055$$

$$P(\text{Shoot} \mid \text{action}) = \frac{4+11}{11+7} = \frac{15}{18} = 0.833$$

$$P(fly \mid action) = \frac{1+1}{11+7} = \frac{2}{18} = 0.11$$

Compute Posterior probabilities :

$P(Comedy \mid D) = P(Comedy) \times P(fast \mid Comedy) \times P(Couple \mid Comedy) \times P(Shoot \mid Comedy) \times P(fly \mid Comedy)$

D = document

$= 0.4 \times 0.125 \times 0.1875 \times 0.0625 \times 0.125$

$= \boxed{0.000073}$

$P(action \mid D) = 0.6 \times 0.166 \times 0.055 \times 0.833$

$= \boxed{0.00456}$

$\because P(action \mid D) > P(Comedy \mid D)$

$\therefore$ The most likely Class for document D is action.

Question 4.3: Solution:

Given document data:

doc	'good'	'poor'	'great'	class
d_1	3	0	3	pos
d_2	0	1	2	pos
d_3	1	3	0	neg
d_4	1	5	2	neg
d_5	0	2	0	neg

Vocabulary = {good, poor, great}

Vocabulary Size = 3

Prior Probabilities:

$$P(pos) = \frac{N_{pos}}{N} \quad \begin{array}{l}(\text{no. of + docs})\\(\text{Total docs})\end{array}$$

$$= \frac{2}{5} = \boxed{0.4}$$

$$P(neg) = \frac{3}{5} = \boxed{0.6}$$

(for multinomial NB)

Likelihoods with Add-1 Smoothing:

$$P(\text{word} \mid \text{Class}) = \frac{\text{no. of times word in class} + 1}{\text{total words in class} + \text{Vocab Size}}$$

(total words in all classes w/o repetition)

$$P(\text{'good'} \mid \text{pos}) = \frac{3+1}{9+3} = \frac{4}{12} = \boxed{0.33}$$

$$P(\text{'poor'} \mid \text{pos}) = \frac{1+1}{9+3} = \frac{2}{12} = \boxed{0.166}$$

$$P(\text{'great'} \mid \text{pos}) = \frac{5+1}{9+3} = \frac{6}{12} = \boxed{0.5}$$

$$P(\text{'good'} \mid \text{neg}) = \frac{2+1}{14+3} = \frac{3}{17} = \boxed{0.176}$$

$$P(\text{'poor'} \mid \text{neg}) = \frac{10+1}{14+3} = \frac{11}{17} = \boxed{0.647}$$

$$P(\text{'great'} \mid neg) = \frac{2+1}{14+3} = \frac{3}{17} = \boxed{0.176}$$

<u>For Binarized NB:</u>

$$P(w \mid Class) = \frac{\text{no. of doc in class containing } w + 1}{\text{Total no. of docs in class} + 2}$$

w ↳ word

$$P(\text{'good'} \mid pos) = \frac{1+1}{2+2} = \frac{2}{4} = \boxed{0.5}$$

$$P(\text{'poor'} \mid pos) = \frac{1+1}{2+2} = \frac{2}{4} = \boxed{0.5}$$

$$P(\text{'great'} \mid pos) = \frac{2+1}{2+2} = \frac{3}{4} = \boxed{0.75}$$

$$P(\text{'good'} \mid neg) = \frac{2+1}{3+2} = \frac{3}{5} = \boxed{0.6}$$

$$P(\text{'poor'} \mid neg) = \frac{3+1}{3+2} = \frac{4}{5} = \boxed{0.8}$$

$$P(\text{'great} \mid neg) = \frac{1+1}{3+2} = \frac{2}{5} = \boxed{0.4}$$

Classify the Sentence:

Using multinomial NB:

P(pos | Sentence) = P(pos) x
P('good' | pos) x
P('good' | pos) x
P('poor' | pos) x
P('great' | pos)

drop words that are not present in either class.

$= 0.4 \times (0.33)^2 \times 0.166 \times 0.5$

$= \boxed{0.00361}$

P(neg | sentence) = $0.6 \times (0.176)^2 \times$
0.647×0.176
$= \boxed{0.00211}$

∵ P(neg | Sentence) > P(pos | Sentence).
∴ multinomial NB would Classify the Sentence as Negative.

Using Binarized NB:

P(pos | sentence) = P(pos) x P('good' | pos) x P('poor' | pos) x P('great' | pos)

multiple Prob for same words are ignored in binary only 1 Prob

$= 0.4 \times 0.5 \times 0.5 \times 0.75 = \boxed{0.075}$

$P(neg \mid Sentence) = P(neg) \times P(\text{'good'} \mid neg) \times P(\text{'poor'} \mid neg) \times P(\text{'great'} \mid neg)$

$= 0.6 \times 0.6 \times 0.8 \times 0.4$

$= \boxed{0.1152}$

$\because P(neg \mid Sentence) > P(pos \mid Sentence)$

$\therefore$ Binary NB would also classify the Sentence as Negative.

Conclusion: Both the models agree on that the given Sentence is a negative Sentence.

CamScanner