"NLP ASSIGNMENT 1"

Date

Q 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	0.25
film	0.08	0.11

PRIOR PROBABILITIES (equal):
$P(pos) = P(neg) = 0.5$

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

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

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

$\Rightarrow P(neg \mid \text{'I always like foreign films'}) = 0.5 \times 0.16 \times 0.06 \times 0.25 \times 0.11$

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

$\because P(pos \mid \text{'I always like foreign films'}) > P(neg \mid \text{'I always like foreign films'})$

$\therefore$ Naive Bayes would classify the sentence as positive.

Q4.2: SOLUTION:-

Step #1: Computing Prior Probabilities

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

$P(\text{Action}) = 3/5 = 0.6$

Comedy Class:-

Words	Count
Fun	3
Couple	2
Love	2
Fly	1
Fast	1

Total words in comedy class = 9

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}
(Total words)

Vocabulary Size : 7

Likelihoods with Add-1 Smoothing:

$$P(\text{words} \mid \text{class}) = \frac{\text{count (word in class)} + 1}{\text{total words in class} + \text{Vocabulary size}}$$

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:

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

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

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

$P(\text{Fly} \mid \text{Action}) = \frac{1+1}{11+7} = \frac{2}{18} = 0.11$

Compute Posterior Probabilities:

$P(\text{Comedy} \mid D) = P(\text{Comedy}) \times P(\text{Fast} \mid \text{Comedy}) \times P(\text{Couple} \mid \text{Comedy}) \times P(\text{Shoot} \mid \text{Comedy}) \times P(\text{Fly} \mid \text{Comedy})$

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

$= 0.000073$

$P(\text{action} | D) = 0.6 \times 0.166 \times 0.055 \times 0.833$
$= 0.00456$

$\because P(\text{action} | D) > P(\text{comedy} | D)$

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

Q4.3: 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:-

$P(pos) = \frac{N_{pos}}{N} = \frac{2}{5} = 0.4$

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

For MULTI-NOMINAL NB:

LIKELIHOODS WITH ADD-1 SMOOTHING:

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

$P(\text{'good'} | pos) = \frac{3+1}{9+3} = \frac{4}{12} = 0.33$ $\quad P(\text{'good'} | neg) = \frac{2+1}{14+3} = \frac{3}{17} = 0.176$

$P(\text{'poor'} | pos) = \frac{1+1}{9+3} = \frac{2}{12} = 0.166$ $\quad P(\text{'poor'} | neg) = \frac{10+1}{14+3} = \frac{11}{17} = 0.647$

$P(\text{'great'} | pos) = \frac{5+1}{9+3} = \frac{6}{12} = 0.5$ $\quad P(\text{'great'} | neg) = \frac{2+1}{14+3} = \frac{3}{17} = 0.176$

For BINARIZED NB:

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

$P(\text{'good'} | pos) = \frac{1+1}{2+2} = \frac{2}{4} = 0.5$ $\quad P(\text{'great'} | pos) = \frac{2+1}{2+2} = \frac{3}{4} = 0.75$

$P(\text{'poor'} | pos) = \frac{1+1}{2+2} = \frac{2}{4} = 0.5$ $\quad P(\text{'good'} | neg) = \frac{2+1}{3+2} = \frac{3}{5} = 0.6$

$P(\text{'great'}|neg) = \frac{1+1}{3+2} = \frac{2}{5} = 0.4$ $\quad P(\text{'poor'}|neg) = \frac{3+1}{3+2} = \frac{4}{5} = 0.8$

Classify the sentence:

Using multinominal NB:

$P(pos|sentence) = P(pos) \times P(\text{'good'}|pos) \times P(\text{'good'}|pos) \times P(\text{'poor'}|pos) \times P(\text{'great'}|pos)$

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

$= 0.00361$

$P(neg|sentence) = 0.6 \times (0.176)^2 \times 0.647 \times 0.176$

$= 0.00211$

$\because P(neg|sentence) < P(pos|sentence)$

$\therefore$ Multinominal NB would classify the sentence as Positive.

Using Binarized NB:

$P(pos|sentence) = P(pos) \times P(\text{'good'}|pos) \times P(\text{'poor'}|pos) \times P(\text{'great'}|pos)$

$= 0.4 \times 0.5 \times 0.5 \times 0.75 = 0.075$

$P(neg|sentence) = P(neg) \times P(\text{'good'}|neg) \times P(\text{'poor'}|neg) \times P(\text{'great'}|neg)$

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

$= 0.1152$

$\because P(neg|sentence) > P(pos|sentence)$

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

Conclusion: Both the models agree on that the given sentence is a Negative sentence.