NAME : NAGESWARARAO CLASS ID - 18

1-a) Bernoulli Naive Bayes model:

$P(x_{francisco} = true \mid class = SFO) = \frac{2}{2} = 1.0$

$P(x_{london} = true \mid class = SFO) = \frac{1}{2} = 0.5$

$P(x_{francisco} = true \mid class = JFK) = \frac{1}{1} = 1.0$

b) Multinomial NB model :-

$P(x = francisco \mid class = SFO) = \frac{4}{14}$

$P(x = london \mid class = SFO) = \frac{1}{14}$

$P(x = francisco \mid class = JFK) = \frac{1}{8}$

c) -i) Bernoulli model : It is not very accurate, because it ignores frequency information

ii) more accurate, because it uses frequency information. However it ignore position information so doesn't distinguish between a city name occuring at the beginning/end of itenary from one occuring in the middle.

d) We will use as a feature the term that occurs in the last position of each document.

2-a) It will never choose a category unless all words in a document were seen for that category for training set. It will rank between classes for which all words were seen similar to two smoother classifier.

b) It will be more likely to choose category for which some/many of the words in the document unseen.

3-a) the precision is $tp/tp+fp$

$$= \frac{3}{2+3} = 3/5$$

The recall is $= \frac{tp}{tp+fn}$

$$= 3/8$$

3-5)

(i) The IR system which always returns no results will have high accuracy for most queries, since the corpus usually contains only a few relevant documents. Documents that are truly relevant are the only ones that will be mistakenly classified as irrelevant, and the accuracy is close to 1. Recall and precision are two different measures that can jointly capture the tradeoff between returning more relevant results and returning fewer irrelevant results.

(ii) There are of course many correct answers. The simple answer is:

Assume document 1 is only relevant document

$A_q = \{1, 2, 3\}$

$B_q = \{3\}$

Both A_q and B_q make 2 mistakes. So they have accuracy 80%.

The precision of A_2 is $\frac{1}{3}$, the precision for B_2 is 0.. since B_2 didn't return any relevant documents. it is no utility.