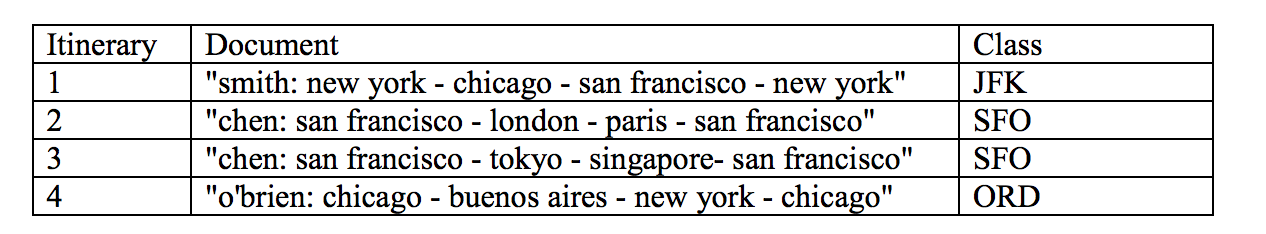
**Knowledge Management and Discovery**

**Problem Set- 6**

**Name : Rakesh Reddy Pallepati**

**Class Id: 20**

1. **Consider the problem of classifying the origination point of passenger travel itineraries. Suppose we have the following training set of travel itineraries:**

****

1. **Assume that we use a Bernoulli (i.e., binary) Naive Bayes model. Compute the following feature probabilities:**

* **P(Xfrancisco=true | Class=SFO)**
* **P(Xlondon=true | Class=SFO)**
* **P(Xfrancisco=true | Class=JFK)**

**A):**

P(Xfrancisco=true | Class=SFO) = P (Class=SFO|Xfrancisco=true)\*PXfrancisco=true)/P(Class=SFO)

= (1\*1/2) / (1/2) = 1

P(Xlondon=true | Class=SFO) = P(Class=SFO|Xlondon=true)\*P(Xlondon=true)/P(Class=SFO)

= 0.5

P(Xfrancisco=true | Class=JFK) = P(Class=JFK|Xfrancisco=true)\*P(Xfrancisco=true)/P(Class=JFK)

= 1.0

1. Assume that we use a multinomial NB model instead. Compute the following probabilities:

* P(X=francisco | Class=SFO)
* P(X=london | Class=SFO)
* P(X=francisco | Class=JFK)

Ans:

P(Xfrancisco|SFO)=4/14

P(X=london|SFO)=1/14

P(Xfrancisco|JFK)=1/8

1. **Consider a standard Naive Bayes classifier trained on the training set and applied to a similar test set. How accurate is this classifier for:**
2. **the Bernoulli model, and**
3. **the multinomial model?**

**Ans:**

Consider a standard Naive Bayes classifier trained on the training set and applied to a similar test set. How accurate is this classifier for:

(i)the Bernoulli model : for given data we can say that Bernoulli model not accurate which ignores term frequency which can be important

(ii)the multinomial model: for given data we can say that multinomial model more accurate because it uses freq information

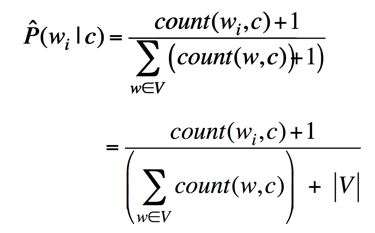
**d.Construct a non-standard feature representation that is 100% accurate for either model**

Ans:P(Xnewyork=true|JFK)=1.0

P(XSan Francisco=true|JFK)=1.0

P(XChicago=true|JFK)=1.0

1. **This problem concerns smoothing Naïve Bayes classifiers. Consider the following formula for Laplace (add-1) smoothing for Naïve Bayes**

****

1. **Suppose we build a Naive Bayes classifier (multinomial or Bernoulli) with no smoothing of the respective P(word | class) probabilities. If a word was unseen in a class, it will thus have a probability of 0. Describe in words the decision procedure of this classifier (emphasizing the effect of the lack of smoothing, and how its decisions will differ from a smoothed Naive Bayes classifier).**

**Ans:**

In section I, I used Bernoulli Naïve Bays model without smoothing while used multinomial with smoothing. So with the reference of that we can say that, if we use without smoothing, for we get some probability 0 and 1 directly. On other hand for multinomial, I used smoothing that why I got out other than zero for that case where conditional probability is zero.

To see why consider the worst case where none of the words in the training sample appear in the test sentence. In this case, under your model we would conclude that the sentence is impossible but it clearly exists creating a contradiction.

Another extreme example is the test sentence "Alex met Steve." where "met" appears several times in the training sample but "Alex" and "Steve" don't. Your model would conclude this statement is very likely which is not true.

1. **Suppose we take a smoothed multinomial classifier and double the amount of smoothing (e.g., for a variant of “add 1 smoothing”, add 2 to each count, and add to the denominator 2k, where k is the number of samples). What qualitative effect will this have on decisions of the classifier?**

Ans: Consider an example of when a word is repeated number of times then the denominator will be much higher and it will reduce the probability for that word. And for the word which appears less time has lower denominator, so probability is higher. This means the result we get is not correct.

**III) An IR system returns 3 relevant documents, and 2 irrelevant documents. There are a total of 8 relevant documents in the collection.**

1. **What is the precision of the system on this search, and what is its recall?**
2. **Instead of using recall/precision for evaluating IR systems, we could use accuracy of classification. Consider a classifier that classifies documents as being either relevant or non-relevant. The accuracy of a classifier that makes c correct decisions and i incorrect decisions is defined as: c/(c+i).**
3. **Why do the recall and precision measures reflect the utility (i.e., quality or usefulness) of an IR system better than accuracy does?**
4. **Suppose that we have a collection of 10 documents, and two different boolean retrieval systems A and B. Give an example of two result sets, Aq and Bq, assumed to have been returned by the system in response to a query q, constructed such that Aq has clearly higher utility and a better score for precision than Bq, but such that Aq and Bq have the same scores on accuracy.**

a)

Given, total 5 calls

3 relevant documents means recalled successfully and two are irrelevant means not recalled.

So precision for the system is 3/5= 0.6 (60%)

Its recall is simply 3/8 as it retrieves 3 correct out of 8. b)

i)

The accuracy given is defined by c/(c+i).

Take an example for spam detecting system.

And we got below results.

Classified positive Classified negative

Positive class 0 (TP) 100 (FN)

Negative class 0 (FP) 100 (TN)

For this output the accuracy will be 100/200 = 50% which is higher but actually the result is total different. The actual precision of the system is 0%.

We need recall and precision because, for accuracy there is no concept false positive and false negative.

ii) Doc1={1,2,3}

Doc2={3}

Here doc1&2 have same accuracy 80%

Doc1=Prec=1/3

Doc2=0

Doc2 is of no use because of no relevant docs