

## Written Assignment 2

### CS 510 TOP: Information Retrieval on the Internet

Assigned: April 20, 2010

Due: April 27, 2010

Possible points: 8

1. Suppose you have a collection of 100 documents, of which you know 20 are **relevant** to a specific query. Consider the following results from a search of this collection using this query, which retrieved a total of 30 documents.

Rank    Relevant?

Rank	Relevant?
Doc1	Y
Doc2	N
Doc3	Y
Doc4	Y
Doc5	N
Doc6	Y
Doc7	Y
Doc8	Y
Doc9	Y
Doc10	Y
Doc11	Y
Doc12	Y
Doc13	Y
Doc14	N
Doc15	Y
Doc16	N
Doc17	Y
Doc18	Y
Doc19	Y
Doc20	Y
Doc21	Y
Doc22	Y
Doc23	Y
Doc24	N
Doc25	N
Doc26	N
Doc27	N
Doc28	N
Doc29	N
Doc30	Y

- (a) What is the precision of these search results?
- (b) What is the recall of these search results?
- (c) Give a table (similar to table 8.1 in the textbook) with the eleven-point interpolated precision for these search results.
- (d) Give a precision/recall plot using the data from (c).
- (e) Give the R-precision of these search results.

2. (a) Create a table like the one you created for 1(c), but here for the recall (i.e., *sensitivity*) versus false positive rate (i.e.,  $(1 - \textit{specificity})$ ) of these search results (see Figure 8.4).
- (b) Plot a ROC curve using the table from 2(a).
- (c) Comment on any major differences you note between your precision/recall and ROC curves for this data.
3. Suppose you have a text classification problem in which documents  $\mathbf{x}$  are classified on the basis of three binary attributes:  $\mathbf{x} = (a_1, a_2, a_3)$ , where  $a_i \in \{0,1\}$ . Given the training set below, show how a naïve Bayes classification algorithm would classify the new instance  $\mathbf{x} = (0, 0, 0)$ , using Laplace smoothing. Give all necessary conditional probability estimates. (You don't need to use the logarithmic form of Naïve Bayes.)

**Training set:**

$\mathbf{x}_1 = (1, 0, 0)$ , class = +

$\mathbf{x}_2 = (0, 1, 1)$ , class = +

$\mathbf{x}_3 = (1, 1, 0)$ , class = -

$\mathbf{x}_4 = (0, 0, 1)$ , class = -

$\mathbf{x}_5 = (1, 1, 1)$ , class = +