CSE 6240 - Spring 2015
Web Search & Text Mining

Homework 3

01/26/2015

Due: 02/01/2015 23:59

1. Rocchio's algorithm (25pt)

a. In Rocchio's algorithm, what weight setting for α/β/γ does a "Find pages like this one" search correspond to?

b. Under what conditions would the modified query q_m in In Rocchio's algorithm be the same as the original query q_0 ? In all other cases, is q_m closer than q0 to the centroid of the relevant documents?

2. Relevance feedback (25pt)

a. Give three reasons why relevance feedback has been little used in web search.

b. Why is positive feedback likely to be more useful than negative feedback to an IR system? Why might only using one non-relevant document be more effective than using several?ta

3. Boosting (25pt)

What's the relation between Adaboost and Boosting? What's the difference between Adaboost and Gradient Boosting? Please explain briefly.

4. Adaboost (25pt)

Below is a 2-classify training set.

Sample	1	2	3	4	5	6	7	8	9	10
X	0	1	2	3	4	5	6	7	8	9
Y	1	1	1	-1	-1	-1	1	1	1	-1

Assume we have a weak classifier G. G uses a threshold v to predict Y, which is generated by x<v or x>v. G uses the v that minimize the error rate for the classification. Using the definition and equations in the slides, show how you get a strong classifier by Adaboost.