# Your Name **Chao Zhang**

# Your Andrew ID **chaozhan**

# Homework 5

**0. Statement of Assurance**

You must certify that all of the material that you submit is original work that was done only by you. If your report does not have this statement, it will not be graded.

**I certify that all of the material that I submit is original work that was done only by myself.**

# 1. Performance evaluation (60%)

## 1.1 P@10, NDCG@10 and MAP (10%)

Write down the formula of each metric, and give one example for each of them to show that single metric is not enough to guarantee the ranking quality.

, whereis the number of relevant documents in top N results.

, whereis the number of kinds of queries, is the number of returned documents, is the precision at cut-off in the returned documents for query and is the change in recall from documentsto.

, where is the ranked list given query , is the document with rank ,is graded relevance and is the ideal ranked list given query .

## 1.2 Performance and time cost table (20%)

Fill in the table with your experiment results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C | Logistic Regression | | | SVM | | |
| P@10 | NDCG@10 | MAP | P@10 | NDCG@10 | MAP |
| 0.0001 | 0.15 | 0.26976 | 0.20815 | 0.09 | 0.12107 | 0.06547 |
| 0.001 | 0.15 | 0.26976 | 0.20815 | 0.07 | 0.11740 | 0.08386 |
| 0.01 | 0.15 | 0.26976 | 0.20812 | 0.11 | 0.19305 | 0.16331 |
| 0.1 | 0.15 | 0.26629 | 0.20567 | 0.14 | 0.24868 | 0.18974 |
| 1 | 0.13 | 0.24226 | 0.19944 | 0.16 | 0.28452 | 0.21766 |
| 10 | 0.12 | 0.22163 | 0.18324 | 0.18 | 0.33832 | 0.26437 |
| 50 | 0.11 | 0.19422 | 0.16428 | 0.15 | 0.31449 | 0.25966 |
| 100 | 0.1 | 0.16804 | 0.13626 | 0.15 | 0.31312 | 0.26007 |

|  |  |  |
| --- | --- | --- |
| C | Time | |
| Logistic Regression (s) | SVM (s) |
| 0.0001 | 173.81042099 | 860.958234980 |
| 0.001 | 179.506263971 | 642.603888713 |
| 0.01 | 181.110648155 | 445.309245349 |
| 0.1 | 127.192248821 | 228.616918354 |
| 1 | 99.2953209877 | 53.788590546 |
| 10 | 42.9407329559 | 30.562038442 |
| 50 | 18.1144480705 | 34.281250796 |
| 100 | 12.60152792931 | 35.709726224 |

## 1.3 Plots (10%)

Plots graphs for each metric of Logistic Regression and SVM. You should have six graphs in total.

## 1.4 Analysis (20%)

Analysis based on your experiment results (performance and time), and some conclusions (What did you learn from the results? What are the advantages and disadvantages of using Logistic Regression and SVM? etc)

# 2. Feature Design (25%)

Add three new features that you think might work, and test it. Give some analysis and conclusions based on the experiment results.

# 3. The software implementation (5%)

Description of your code including any data structures used, design considerations etc. In addition, please describe any changes you made to your system after HW4.