ECE-467 PROJECT 1
WRITE-UP

Derek Lee
Professor Sable

March 14, 2021

1 Instructions

The program uses the NLTK and Punkt libraries. It uses Python 3.8.5.

Run the program with the following command:

```
python3 -O main.py
```

2 Description

Which basic machine learning method did you use?
I used Rocchio/TF*IDF.

How does your system tokenize training and test files?
My system uses the NLTK library to tokenize the input.

What weighting scheme, if any, is used for tokens?
My system uses TF*IDF weighting.

Which optional parameters or features did you experiment with (e.g., possibilities might include case sensitivity, POS tagging, stop lists, etc.)? Which parameters or features made a significant difference, and how are they set in your final system?
I experimented with BM25 weighting instead of TF*IDF. However, this did not lead to an improvement in the performance of my system, so I went back to TF*IDF.

How did you evaluate your system's performance for the second and third data sets?
I split the training set with a 80-20 split, resulting in a new training set and a validation set. I trained the system on the new training set and evaluated the performance on the validation set.

You may include any additional information that you wish.
During testing, I ignored any tokens that were not seen during training.