Grading for Assignment 1

Grading rubric from the assignment description

10% of your grade (10 points) is based on the report. Make sure to answer every question. The other 90% is based on the performance of your code compared to the official solution written by the teaching assistants following the directions above. If your performance is the same or better then you receive full credit. Otherwise, your score is proportional to your relative performance:

90 points * (1 – (performance_of_solution – your_performance) / performance_of_solution)

Breakdown of points for report

- Q1 = 2 points.
- Q2 = 2 points.
- Q3 = 4 points. Must describe an enhancement beyond unigrams as features, add-one smoothing and ignoring unknown words in testing.
- Q4 = 2 points.

Partial credit/regrading

According to the rubric, 90 points are assigned based solely on performance of your code on Vocareum on the unseen test data. The TA solution achieved a weighted F1-score of 98% on the development data. If your code did not achieve the same performance, this could explain low performance on the test data. You should be sure that any code you are reusing for assignment 2 did not cause problems in assignment 1. The TA solution trained on the training and development data achieved a weighted F1-score of 98% on the test data.

After running all the submissions, we noticed two major problems and allow partial credit for them.

Case 1: file reading and writing was not using latin1 encoding which caused the program to crash on submission. We have fixed all these errors, rerun the code to calculate a new score out of 90, and deducted 30 points from this score with the lowest possible score being 0. The 30 point deduction reflects the fact that you would have seen the error when you submitted the code (and should have fixed it) and is better than losing 90 points automatically.

Case 2: did not handle the test data correctly (i.e., did not follow the instructions in the assignment which are repeated below) and the code crashed during grading or returned no results:

> you should not make any assumptions about the structure of the directory. Instead, you should search the directory for files with the extension ".txt"…classify each ".txt" file in the data directory as "ham" or "spam"

The TAs may not have been able to fix your code in which case you will have to meet with one of the TAs and fix the code. The TAs will then rerun your code to calculate a new score out of 90, and <u>deduct 20 points from this score</u> with the lowest possible score being 0. This is better than losing 90 points automatically. It is possible that a submission could fall into case 1 and case 2 and would end up with a 50 point deduction.

Currently, no other partial credit is allowed and all other cases will be graded according to the original rubric.

Who to talk to: <u>Talk to a TA</u>. The instructors will not review your code. Meet with a TA during their office hours, or make an appointment to speak with them. In some cases, your grade already reflects regrading (see the comments), and there is no need to talk with a TA.

Timeline: <u>Must be done before November 5th</u>. You cannot request an adjustment afterwards.