Final Project

ITCS 6156 / ITCS 8156 – Fall 2015

For this project, you have been assigned a real-world machine learning problem. You will implement a solution and discuss the results of your efforts during an in-class presentation. Each complete project will consist of (1) implementation, (2) experimentation, and (3) discussion.

Competition

You have selected (or been assigned to) one of three different projects hosted on Kaggle. Once invited, you should go to https://inclass.kaggle.com/ and find your respective project. Once registered, be sure to pay special attention to the rules of the competition and the format of your results. You will submit your results to Kaggle as 2-column CSV file. For all of the examples in the testing set, the first column is the test ID and the second column depends on the competition. As the competition progresses, you will be able to see a "public" leaderboard. These results are a ranking based on a randomly selected subset of the solutions that you provide (20% of the testing data). Your *real* score (on 100% of the testing data) is kept on a "private" leaderboard that won't be made available until the end of the competition.

The competition **closes on December 13th**, so be sure you have at least one valid attempt submitted to Kaggle. Overall results will be announced during the final project presentations.

The goal of this project is for you to apply machine learning techniques to real-world problems. It is expected that you will use both material from class and outside resources (e.g., papers, code libraries) as part of your work. The data for these problems comes from UNCC researchers. Any attempt to obtain or access the ground truth predictions for the test examples will treated as an academic integrity violation and dealt with harshly.

Project Milestones

To ensure progress toward a complete, reasonable final project, there will be graded intermediate milestones. Credit will only be given for milestone submissions submitted **on time via Moodle**. Do not wait until the last minute to make uploads to Moodle, as it will likely be slow at the end of the semester. No exceptions will be made for the deadlines.

ms0: Project Checkpoint [20 pts]

For this milestone, you must download the data for your project, make at least one submission to Kaggle, and submit a short description of what you have attempted so far and your plan for the project. The majority of the credit on this assignment will be based on my assessment of the reasonableness of your initial attempts and future plans.

ms1: Project Abstract [20 pts]

You will submit a 200 – 300 word abstract describing your problem, solution, and (expected) results. The abstract should provide enough information for a student in the course to understand the problem and your approach. The attached document (from Tufts University) provides tips and examples for writing

abstracts. This milestone will be graded how accurately it reflects the material presented, clarity, and comprehensiveness.

ms2: Project Code [50 pts]

You will submit your project code and a README describing how to run the code and recreate the results that you present. Your code will be graded based on the Code Submission Guidelines described in the course syllabus and how closely it matches your presentation (ms3). You are allowed to use external helper functions and libraries (giving appropriate credit, of course). This is a programming assignment, so using a GUI-based tool (e.g., Weka) will not get you much credit. To reward better-performing algorithms, the maximum points that you can earn for this section will be based on the final ranking of your algorithm, based on the private leaderboard at the end of the competition.

ms3: Project Presentation [50 pts]

You will present a "flash talk" on 12/16 during the assigned final exam period. Each talk is allotted 4 minutes. The time limits will be strictly enforced. Your grade for this milestone will be mainly determined by your classmates, so you'll want to make sure that your presentation is clear, **well-rehearsed**, and describes your approach and results effectively. These few minutes will be your only chance to present the entirety of your final project. It is highly recommended that you rehearse your presentation (multiple times) and put effort into delivering a polished talk. Your score for this milestone could be negatively affected by presenting work different than the proposal or misrepresenting your work and/or results.

To minimize wasted time between presentations, we will run **all** of the presentations on the classroom PC, so you will need to submit your presentation slides via Moodle by the deadline. You will not be allowed to present from your own device or using a presentation other than the one uploaded to Moodle. Be sure to test your presentation on a UNCC Windows 7 image machine (class podium or 120 lab) for codec compatibility and/or to make certain everything looks right.
