MSAN 621 Introduction to Machine Learning Fall 2016 Module 2

Project

The group project for the module is to apply what you know about machine learning to a real dataset or real-world problem. You and your group will be responsible for defining a problem, acquiring data to address the problem, designing and implementing a solution to the problem and finally, presenting your solution to your peers.

Data

You and your group may choose to use an existing dataset, augment an existing dataset or create your own dataset. If you choose an existing dataset, ensure that it is composed from actual data — i.e. it may not contain aggregated data or other data items which can be used as features for your system. You may get this data from any source as long as the data is freely and legally available to the public, even if it requires registration. Alongside many other sources, Kaggle, SF Open Data and data.gov are all reasonable places to acquire data for this project.

Submission

You and your group will submit four items for this assignment:

- A written report with project details about the project. This report should include a description of the project, link(s) to the data source(s), a description of the algorithm(s) attempted and an evaluation of the efficacy of your solution.
- Code and scripts used in implementing the solution.
- An in-group peer evaluation, to be completed online after your group presentation.
- An oral presentation of the project, detailed below.

In addition to these, two project proposals are due — one initial, one final — to be submitted by one member of the group on the Slack channel for the course. These proposals may become the initial portion of the written report. All non-code, non-review submissions must contain the name of the group and the names of the individuals in the group.

Presentation

Your group will present your project in class at one of the times reserved for group presentations. At this presentation, all group members must describe the portions of the projects for which they are responsible. The group may use presentation software (Powerpoint, Google Slides, etc.) to describe the project to the class and to zero or more invited guests. In total, the group presentation should answer the following questions:

1. What was the problem?
2. What hypotheses did you develop about solving the problem?
3. What data and algorithms were used, and why?
4. What is your evaluation of your solution to the problem?
5. What would be your next step for solutions to this problem?

Your group presentation will be 20 minutes total, with 5 minutes reserved for questions. You may elect to take questions during the presentation or at the conclusion of the presentation.

Important Dates

The following table lists all dates and deliverables for the group project:

Date	**Deliverable**
28th October	Form groups formed by instructor
1st November	Initial project proposals due from groups
7th November	Review of project proposals due from instructor
14th November	Final project proposals due from groups
6th December	Group Presentations (round 1)
8th December	Group Presentations (round 2)

Grading

Your group receives an overall grade for the project and for the presentation. This grade is determined as follows:

- 40% = Implementation
    - (20%) Correctness of code (functionality, completeness, etc.)
    - (10%) Readability of code (functions, variable names, comments, etc.)
    - (10%) Reproducibility (with directions, code can be run to reproduce results)
- 30% = Written Report and 30% Presentation
    - (9%) Completeness (eg. does the presentation cover everything required for understanding the project?)
    - (9%) Clarity (approach is clear to peers)
    - (9%) Organization (presentation flow is smooth; report sections are clear and reasonable)
    - (3%) Professionalism

Individual grades will be assigned according to the review from your peers as an adjustment to the group grade.