CMSC 422 - Introduction to Machine Learning Final Project

Fall Semester 2022

Instructor

Mohammad Nayeem Teli, nayeem@cs.umd.edu, Section 0101.

1 Summary

CMSC422 students will turn in a final project that will walk users through the application of one or more algorithms that we have learned in this course. This is a group project (three to five students per group) focusing on deeper understanding of an advanced machine learning problem. Students may choose an application area and dataset(s) that are of interest to them; please feel free to be creative about this! For some ideas and possible data sources, see the following sources:

- https://www.data.gov
- https://cloud.google.com/bigquery/public-data/
- https://www.kaggle.com/datasets
- https://aws.amazon.com/public-datasets/
- https://archive.ics.uci.edu/ml/datasets.php
- https://guides.library.cmu.edu/machine-learning/datasets
- https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html
- https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html
- http://xviewdataset.org/#dataset
- https://image-net.org/
- https://github.com/codebrainz/color-names/blob/master/output/colors.csv

You can find a lot more, here, https://tinyurl.com/t45jk8pv, or work something of your own. You may write all your code in a Jupyter notebook. The notebook should have a mix of markdown prose (explanations) and Python code. Recommended, but not required, you may upload your code as a Github statically-hosted Page.

1.1 What is due

- 1. Project presentation, due on Dec. 6 and 8, 2022 in class. (The order of the presentations will be posted on Piazza).
- 2. Jupyter Notebook or Github page (hosting the notebook), due on Thursday, December 15, 2022.
- 3. Project report due on Thursday, December 15, 2022

The presentation will be in class (last two lectures). Each group may present for 5 to 7 minutes. You are expected to upload your project slides as a pdf file to ELMS on **December 8**, **2022**. The report should have at least 1500 words of prose, along with appropriate documentation, visualization, and links to any external information that might help the reader. You are welcome to do this project individually as well.

1.2 Github Pages

GitHub provides a service called Pages (https://pages.github.com/) that provides website hosting functionality backed by a GitHub-based git repository. We would recommend to host your final project on a GitHub Pages project site. To do this, you will need to:

- 1. Create a GitHub account (or use the one you already have) with username <username>.
- 2. Create a git repository titled **username**.github.io; make sure username is the same as whatever you chose for your global GitHub account.
- 3. Create a project within this repository. This is where you'll dump your iPython Notebook file and an HTML export of that Notebook file.

These instructions are also given directly on the front page of https://pages.github.com/; following those instructions should be fine!

If you decide to host your. project on Github, the deliverable to the CMSC422 staff will then be a single URL instead of the Jupyter Notebook pointing to this publicly-hosted GitHub Pages-backed website. It is due by **11:59PM on Thursday, Dec. 15th**. We will not (*cannot*) accept late assignments. The slides and the report (pdf file) still need to be uploaded to ELMS.

Please make sure to include your name (and the names of all group members) at the top of your deliverable, after the title.

2 Grading

We will assign a numeric score between 1 and 10 for each of the following six dimensions:

- 1. Motivation. Does the project make the reader believe the topic is relevant or important (i) in general and (ii) with respect to Machine Learning?
- 2. Understanding. After reading through the project material, does an uninformed reader feel informed about the topic? Would a reader who already knew about the topic feel like s/he learned more about it?
- 3. Other resources. Does the project link out to other resources (on the web, in books, etc) that would give a lagging reader additional help on specific topics, or an advanced reader the ability to dive more deeply into a specific application area or technique?
- 4. Prose. Does the prose portion of the report actually add to the content of the deliverable?
- 5. **Code.** Is the code well written, well documented, reproducible, and does it help the reader understand the project? Does it give good examples of specific techniques?

6. **Subjective evaluation.** If somebody linked to this project from, say, Hacker News, would people actually read through the entire thing?

Dimension	Max Points
Motivation	10
Understanding	10
Further Resources	10
Prose	10
Code	10
Subjective Evaluation	10
Total	60