Project 2

In this project you will analyze a dataset much as you did in Project 1.

However for this project, instead of a blog post, you will write a short accompanying paper.

Below I describe the sections that the paper should have and what the content of each section should be.

The overall length of the paper should be about 3,4 or 5 pages.

You will choose just one dataset (pick either classification or regression).

But in this case please try and compare several different methods (see the Methods section described below).

For examples of papers like this you can browse through these (the reports not the papers):

http://cs229.stanford.edu/proj2019aut/

Those papers are all extremely good -- yours doesn't have to be that good.

If your paper is also very good then great, but considering time limitations, Covid, etc, you can get an A simply by ticking all the boxes outlined below.

Paper sections

- Abstract

Briefly describe what you did -- which dataset you analyzed, which methods you tried, and what the results were.

- Introduction

Explain to a non-expert what problem you solve in the paper. Give a little background on the kind of task you are doing -- what methods, what data. Cite previous work on the general problem (eg optical character recognition).

Describe in brief what you will show in the remaining sections of the paper.

- Dataset

Give particulars about what your dataset is, the number of instances, the number of columns, and possibly some simple statistics.

Explain the meaning of the dataset and the target, including the meanings of the columns.

If appropriate cite previous work on this dataset (most UCI ML Repo entries provide a reference you can cite).

- Methods

Explain the techniques you will use to analyze the data.

For example if you are doing classification, describe what classifiers you will use.

Say a little bit about how they work.

If you are doing any feature transformations, explain those.

- Results

Present the results of your work.

The central item here should be a table showing the relative performance of the different models you tried.

The table should also be described in prose.

You should mention training (or cross-validation) performance as well as test performance.

- Conclusion

Review what you did and what its significance is.

Draw conclusions about what you learned from your results.

For example which method worked best. Why might it have been the best?

Allude to some further experiments you would like to try in the future.

- Bibliography

Provide a bibliography for the references you cited. This can include actual papers, but also blog posts.