Final Project

CSCI 297: Introduction to Machine Learning

Due: Saturday November 21st @ 11:59 PM

Professor Watson

This is your final project where you will be implementing a machine learning classifier on a dataset that you curated. You have already extracted the necessary data to run your algorithm, now you need to perform all the steps in applying a machine learning based solution.

In your project, you show evidence of data extraction, data imputation, feature selection, feature extraction, feature scaling, and EDA. These steps will comprise the first part of your grade for the final project. An explanation of these portion should be captured in your documentation and research paper.

The next portion of your project involves the model creation and selection phase. In this portion of your project you should implement a variety of learning algorithms to analyze your data. This portion of your project should contain the implementation of different algorithms, different generalization and regularization techniques, and the exploration of optimal hyperparameters. You may also try some ensemble methods if you believe these would be beneficial. Again, this portion of your project must be captured in the documentation and your short research paper.

The last portion of your project is the testing and evaluation of the model. This involves an analysis of training time, training effectiveness, Occam’s Razor consideration, over- and underfitting analysis and the use of different evaluation metrics. There may be some cross over between the evaluation and hyperparameter search as you try to find the optimal hyperparameters given a validation set. You should also present an argument in your research paper about how your ML approach is generalizable and could be applied to unseen data.

In addition to comprehensive documentation of your code, decisions made about your ML approach and what strategies were attempted but were unsuccessful, you will need to write a short research paper on your project. This research paper should have the following sections: introduction, related works, methodology, evaluation, threats to validity and conclusion/future work. A short paper means it should be no longer than 6 pages. The paper should include figures, which help the reader to better understand your methodology and the results of your evaluation.

All of these files should be turned in on canvas/GitHub. If any of you believe your project could be extended into publishable research, please contact me to see if this is possible. Your project should consist of the data, any python files, any documentation files, and the research paper. I have included a template for your research paper on canvas.

Good Luck!