CS 6140: Machine Learning — Fall 2021— Paul Hand

Project Planning Document
Due: Monday November 22, 2021 at 11:59 PM Eastern time via Gradescope.

Names: [Put Your Name(s) Here]

For your final project, you will obtain a dataset, select multiple machine learning models, train the models, and evaluate the performance of the models. You may elect to reproduce some of the results from a scientific paper, but you must code up some aspect of dataset, model, or training yourself. You may use standard Deep Learning frameworks (e.g. PyTorch, TensorFlow, etc.). You may use code that is available on the internet as building blocks. You may run your algorithm in a slightly different context. You must train more than one machine learning model and compare the performance of those models. You are encouraged (but are not required to) train models that we have not discussed in class.

THE FINAL PROJECT IS DUE at 11:59 PM on WEDNESDAY DECEMBER 15, 2021 on Gradescope.

You will write up a short (at most 3 pages) report detailing: the dataset you are using and any data processing you have done, the models you are studying, the details of training the models, and the results of the evaluation. Please use the NeurIPS Style files for your report.

You may work in groups of up to 3 people. You may work alone.

If you want some ideas of projects, here are some ideas. You do not need to select one of these papers.

- Train several handwritten digit classifiers from the table at this website.

- Implement one of the chapters of the Mattmann book.

- Find a Kaggle dataset that you find interesting and train multiple models for it.

- Collect data from your own life and train two predictors that you could use.

- Train a neural network to remove additive noise from images. You can construct a dataset consisting of clean images and noisy images that you construct.

- Create a synthetic dataset and evaluate the k-means and k-means++ algorithms

- Create a synthetic high dimensional dataset and show that k nearest neighbors fails while another classification method succeeds.

- Create a synthetic dataset and evaluate how successful cross-validation is at estimating test error.

- Reproduce aspects of Figure 1 of Understanding Deep Learning Requires Rethinking Generalization

Question 1. *Project Planning*

1. Provide a summary of the goal of your project. If you are replicating part of a paper, include a link to the paper here.

   Response:

2. What dataset will you use?

   Response:

3. What models will you train? You need to have more than one.

   Response:

4. What do you think will be most difficult about training the models?

   Response:

5. How will you evaluate the models?

   Response: