EE 660 **Overview of Final Project** Posted: Sat., 10/10/2020
Jenkins Also see timeline information on page 2.

Introduction

For this project you will pick your own topic and design your project. You are encouraged to pick a topic (or dataset) of interest to you, and that is appropriate for a machine learning class project.

Types of Projects

There are two overall types of projects; you may choose either one for your project.

(1) Type 1 project. Solve a machine learning problem by implementing a machine learning system of your own design, that uses real-world data. For this, you will choose one (or more) set(s) of real-world data, and define the goals of your project. For example, the goal of your project might be to use regression or classification techniques to predict the output attribute *y* as well as possible. You could additionally include other goals, such as understanding what the limitations in your final system are caused by; investigating the attributes that are most predictive, and assessing why; etc. You will typically have other issues to address as well, such as number of data points *N* not being ideal, missing or noisy data, imbalance of data set, categorical feature values, preprocessing steps, etc. See the "Dataset Tips" document for suggestions of where to find datasets, and criteria for sifting through them to find one appropriate for a class project.

(2) Type 2 project. Perform one or more experiments in machine learning. The experiments would typically use synthetic data, so that the data can be controlled and varied in various ways; synthetic data also allows you to generate "unknowns" to numerically estimate the out-of-sample error directly. It might also be applied to real-world data to assess the effects of realistic data.

This would typically also involve some theory – either to predict what would happen, or to help interpret the results of what did happen. Experimental work would typically have a statement of what will be learned from the experimental results, or a prediction of what is expected; and explanations and interpretation (after the experiment) based on some theory and intuition. Or, a project might start with a theoretical component that develops some predictions, and then run some numerical experiments to test them.

A good example of an experiment is Sec. 4.1.2 of AML, especially Exercise 4.2, including the results shown in Fig. 4.3 and some of its interpretation.

Suggestion: If you're not sure what you want to do, you can try the following.

(1) For a Type 1 project, start by finding a dataset that you're interested in, and develop a project and goals based on that data. Or, you can also browse through Kaggle competitions to get an idea of what kinds of topics could constitute a project.

(2) For a Type 2 project, you can choose some aspect of class material you find interesting, and pose some questions of how some variables would depend on others; especially where it isn't obvious, where we haven't given examples that show the dependence, or where you can think of a lot more to try than in the examples we covered in class.

Individuals or teams: You may do your own individual project, or you may work in a team of 2 students. Teams of 3 students may be allowed in cases that clearly warrant it. Your project will be graded accordingly; that is, 2 students should accomplish about twice the work of one student (or solve a problem that is an appropriate factor more difficult).

Your submissions and timeline:

You will prepare and submit a **project proposal** (as your Homework 6); more detail and a project proposal form will be posted when the full Project Assignment is posted, and your proposal will be due 7-10 days later.

Your **Final Project Report** will be written and will be typeset; it will describe your approach and results. It, and your computer code, will be due on Friday, 12/3/2020.