Welcome to this course, Machine Learning for Accounting with Python. I’m Linden Lu and I’m the instructor of this course.

As indicated by the course title, this course is about machine learning. You will learn various machine learning algorithms and how to apply them to business related datasets with python.

It is very important to always have the big picture in mind so that you are not overwhelmed by the algorithms or the python coding details. So let’s first look at the whole data analysis process.

(slide1)This is one of the most popular data analysis framework, CRISP-DM, or cross industry standard process for data mining. I’m not going to explain this framework in detail since you should have already seen this in the course Accounting Data analytics with python, which is the prerequisite for this course.

(slide2)Accounting Data analytics with python covers the second and the third steps in the framework, data understanding and data preparation. (slide3)This course covers the next two steps, modeling and model evaluation.

(slide4)Module 1 of this course introduces machine learning in general. You will also learn data pre-processing, which is an important part of modeling.

(slide5)Modules 2 and 3 introduce six machine learning algorithms. A machine learning algorithm is a sequence of rules and instructions to solve a class of problems. Once an algorithm is implemented with a programming language, which is python in this course, it becomes a model. Modules 2 and 3 explain the algorithms and also demonstrate how to construct and train the machine learning models with python code.

(slide6)Module 4 introduces the model evaluation metrics.

The purpose of model evaluation is to improve the model.

So, following model evaluation,(slide7) In module 5, we introduce model optimization, which includes feature selection and model selection. Feature selection helps you select most relative features from a dataset, and model selection determines the best hyperparameters for a machine learning model with the help of cross-validation.

(slide8)Module 6 introduces a special application of machine learning, text analysis, which extracts insights from text documents like SEC filings and customer reviews.

(face)All machine learning algorithms introduced in the first 6 modules are supervised learning, where the models are trained with data whose true outcome or label is available. A trained model develops a map between the income and the outcome of the data, then predicts unseen data with this map.

(slide9)Module 7 introduces an unsupervised learning, clustering, which is designed to extract structures from datasets without true outcome or label. One unsupervised learning example is Customer Segmentation, which divides customers into different groups based on certain characteristics like gender, age, and purchase pattern. We will introduce two clustering algorithms, k-means and DBSCAN.

(slide10)Module 8 introduces a special kind of data, time series data, which is a dataset indexed by time. Hourly weather data and daily stock quote data are two examples of time series data. We will introduce basic date and time manipulation with python. We will also briefly introduce some basic time series analysis technics.

(slide11)This course runs on the same platform as the one used in the prerequisite. Both lesson materials and assignments are in jupyter notebooks. We provide you a jupyter notebook server on coursera, but it’s highly recommended that you install the server on your own machine. The easiest way to have your own notebook server is through anaconda.

(slide12)Now let’s get back to the big picture. After you finish this course, you will be able to perform the entire data analysis process on business related datasets with python. (slide13)In this process, I personally think that the most important steps are business understanding and data understanding because they provide the foundation of a solid analysis. (slide14)Data preparation is the most tedious and time-consuming step, and most coding efforts are dedicated to this step.

(slide15)The two steps covered by this course, Modeling and model evaluation, are the fun part. In these two steps, the python coding is relatively simple since it follows some standard steps, but you can extract exciting insights from the dataset.

(slide16 or face)So, I hope you enjoy this course.