GSCM 410/510

Machine Learning Applications for Supply Chain Management

Online, Summer 2021, 8 weeks

Contact

Professor: David W. Gerbing, Ph.D.

Office: KMC 550D

Instructor Email: gerbing@pdx.edu

Phone: 725-4767 (email works better)

Office Hours: **Thursday 5pm, Sunday online at 5 pm** (before weekly homework is due at
     midnight), and by appointment.
     Each office hour serves as a regularly scheduled video conference.

Time and Classroom: Online Class Website: d2l.pdx.edu

Video Conferences: D2L —> People —> Zoom Meetings, office hours links

Course Introduction

Prerequisites. This course presents all needed statistical and programming concepts from the beginning, so no formal requirement. However, the more background with the following two topics, the less work and time needed to understand the course material.

- Statistics, particularly linear regression analysis (e.g., GSCM 451 or 571, ISQA 516)

- Scripting data analysis language, R or Python

If you have no memory of regression analysis from your required statistics or other courses, and if your data analysis skills are limited to Excel, then be prepared to spend more time than you usually would for a 4-credit course. This emphasis is particularly true given that the summer term compresses the usual 10 weeks of a quarter into 8 weeks.

If you do not like learning statistics, are somewhat intimidated by the computer, do not have a data analysis background beyond Excel, enjoy the beach, work at least one job, and are taking several other summer courses, this course is not for you. This is *not* a course where we read about general principles of machine learning in society and business, sit around in a circle and share our feelings about the awesomeness of it all, then cram the night before for a multiple-choice test. In this course, we do machine learning analyses and learn some underlying statistical methodology, which requires active involvement throughout the term.

This course serves the needs of our GSCM and Data Analytics graduate students. The course also works for undergraduates who are ready to accept the challenge. But then you are learning real data science developing actual machine learning apps, so naturally requires some effort.

Introduction to Machine Learning. The above qualifications aside, machine learning actually is awesome. One form of machine learning uses the computer to detect patterns from which it "learns" the relationships among a set of variables to generate predictions for a variable of interest, assesses the predictive accuracy of its outputs, and modifies itself accordingly to improve the accuracy of future predictions. The predictions are either numerical or classified into one of two or more groups. The profound effects of machine learning contribute not only to business analysis but increasingly influence society, such as Netflix recommendations, medical diagnostics, facial recognition, and self-driving cars.

This course introduces machine learning with illustrative business applications, including the supply chain. The course focuses more on the techniques and methods than on the statistics behind these methods. We apply the most popular platform for doing machine learning, the Python programming language, with no assumption of previous programming experience.

This course introduces the concept, provides some knowledge of Python related to data analysis, provides a basic understanding of how experts apply machine learning, provides basic skills to do some introductory machine learning, and serves as a basis for learning more in the future. To achieve more mastery in Python programming and machine learning, consider more technical courses such as CS 445/545.

Course Overview

1. *The conceptual basis of machine learning*. Beginning with traditional regression analysis, this course introduces the conceptual framework of machine learning for both supervised and unsupervised learning with a focus on applications. Presentations include the types of analyses appropriate for machine learning, indices of model fit for both prediction of a continuous variable and classification, feature extraction and selection, and the variance-bias tradeoff.

2. *The Python programming language as applied to data science*. The basic concepts of Python for data analysis only are presented, such as for reading data sets, simple manipulation of data sets, and accomplishing machine learning analyses. A class example accompanies each presentation of a machine learning technique, followed by a corresponding homework assignment to implement the machine learning analysis. Python is available for free download or for free use in the cloud.

3. *Focus on applications*. This course applies machine learning to actual data. The basic principles of machine learning generalize to a wide range of topics and applications both for the analysis of business processes and beyond, such as inventory control, minimization of transportation costs, demand forecasting, and much more.

Text: There is nothing extra to purchase as **all needed materials are provided as part of the course**. All conceptual materials are presented in the form of an online, interactive textbook. This material is stored outside of D2L on my PSU website, so it is available to students indefinitely, including future revisions.

Machine Learning Python Templates

The Seven Templates. The Python material for each of the seven weeks of content centers around a template. Each template is working Python code with an explanation of the code and what it accomplishes. Each template is a literal guide of how to conduct a specific type of analysis. No one remembers all the details of the working code. Still, if you follow the template to conduct an analysis, the template guides you through the analysis, and you will be able to reference the templates indefinitely into the future.

The template presented each week transfers directly to the corresponding machine learning homework problem, just as it would transfer to conducting such an analysis on a new problem at work. Of course, adjustments are always made, and an expert would likely make many adjustments for each application. Nonetheless, each template is a sufficient starting point for each analysis.

Content. The weekly template for machine learning is presented in two forms:

1. pdf of the content of the template
2. annotated video of the pdf in which I walk through the template step-by-step

Communication

Use my personal email to discuss personal issues, such as your mother-in-law dropping in for a surprise visit right before an assignment is due and you request an extension. For questions regarding the content of the course, use the corresponding D2L discussion forum so that other students can help answer and all can see the question and answers.

Homework

Weekly Homework [40% of grade]: Weekly homework assignments that include short answer conceptual questions, and at least one machine learning analysis, are assigned and turned in via D2L. Find the homework assignment for each week at the bottom of the corresponding weekly content page. The answers to the questions are due at the end of each week, 11:59 pm on each Sunday night. Solutions are posted early the following week.

Homework Groups: You may work with one or two other students, a maximum of three to a group, to form a homework group and turn in the same homework assignment. If you form a group, assess your group members when the term ends according to their homework contribution. Assign each person a numerical weight, including yourself, where the weights sum to 100. A student's homework scores will be downgraded if other group members assign a much lower participation score than indicated by equal participation.

Final

With an 8-week summer course, there is room for only one test, a comprehensive final. The learning evaluated in the final is conceptual, focused on applying machine learning in terms of

concepts. **No computer coding questions appear on the test**, though the test may include computer conceptual questions.

The Final [55% indicated by of grade] is highly structured. All test questions follow directly from the homework concepts and questions, if not literal copies. The data analysis questions on a test are the same questions from the homework, just different data. I run the analyses and display the relevant output on the test for you. For the homework you run your analyses. For the tests, I provide the output for you so that your task is to answer questions about interpreting the output, the same questions from your homework. The study guide for the tests is the homework problems, all of which are provided with solutions.

The Final consists of potentially multiple-choice and short-answer questions taken closed-book, proctored at your computer with a service called Proctorio. The short-answer questions include the interpretation of computer output. Make sure to go to the Content page on D2L and read more about this proctoring process. The proctoring is based on recording video, audio, and screen capture and restricting activities on your computer, such as having windows open other than the Chrome browser window you use for taking the test. If Proctorio senses suspicious activity such as someone entering the room or you repeatedly looking down at your lap, it provides an automated report to your instructor, who may view the recorded material. Proctorio provides an ADA-compliant user experience, as well as straightforward tools for identifying learner accommodations.

Course Grade

Video Conferences [5% of grade]: Students receive participation credit for two video conferences, one during the first four weeks and one during the second half of the course. Obtain video conference credit by either attending the regularly scheduled video conference office hour, or attend a scheduled video announced at least several days in advance. Just show up when you wish to attend. You are free to participate in as many conferences as you wish, but credit is given for just the two required conferences.

Course Percentage: Your course percentage is calculated from the following weights. Your course percentage directly translates into a letter grade. The minimum guarantees are 90% and above an A- or A, 80% up to 90% a B-, B or B+, and 70% up to 80% a C or C+. The cutoffs may be lowered to your favor, but these are the minimum guarantees.

Graduate students: Take the course as GSCM 510 for graduate credit.
1. You have the choice of an independent project in place of the final. One possibility is a data analysis project with data of interest to you for analysis. The data may be your own or taken from the many available possibilities found on the web, including from provided links. Write a motivation of why the analysis is of interest, do the analysis with Python code using at least two different machine learning algorithms, then summarize and interpret the results.

2. For the 5 weeks of ML content, Weeks 3 through 7, there is one additional homework problem for each assignment. Write most to about a complete page that outlines the application of the technique introduced in an application area of your interest.

<div align="center">

Course Outline
</div>

Week 1: Overview of machine learning, Introduction to Python

Week 2: Python data wrangling and pre-processing

Week 3: Regression analysis as machine learning

Week 4: Multiple regression

Week 5: Classification with logistic regression

Week 6: Classification with decision trees

Week 7: Hierarchical and k-means clustering

Week 8: Review and reflection, Topics that would be covered in an additional course

---