DATA 505: Applied Machine Learning

Jameson Watts, Ph.D.

Spring 2021

E-mail: jwatts@willamette.edu
Office Hours: Click Here for Appointment
Office: 209 Mudd Building

Web: jamesonwatts.com
Class Hours: M 06:00-09:50 p.m.
Class Room: *online*

Course Description

Machine learning is becoming a core component of many modern organizational processes. It is a growing field at the intersection of computer science and statistics focused on finding patterns in data. Prominent applications include personalized recommendations, image processing and speech recognition. This course will focus on the application of existing machine learning libraries to practical problems faced by organizations. Through lectures, cases and programming projects, students will learn how to use machine learning to solve real world problems, run evaluations and interpret their results.

Course Format

This course employs various methods, including formal presentations by the instructor, case discussions, simulations, and in-class activities—the approach used depends largely on the class material for a given week. Active participation is paramount to your success in this course. Students are expected to question, challenge, or clarify the material as it is being presented, and to discuss issues/questions raised by your colleagues and/or the instructor.

Course Objectives

1. Creatively engineer new features to help with model performance
2. Identify and correct for sample bias, association bias, and leakage
3. Implement common machine learning algorithms and interpret the results
4. Use an application interface to run a deep learning model
5. Effectively communicate model results in writing and in person

Course Materials

- Base R, Install from here
- Latest Version of RStudio, Install from here
- Various free resources (links are in the schedule on the syllabus)
- Willamette RStudio Server (tentative): Click here

Graded Items

- Homework Assignments: **25%**
- Midterm Exam: **25%**
- Model Performance (x3): **30%**
- Final Exam: **10%**
- Group Presentations: **10%**

Grade Distribution

Percentage	Grade
> 95.00	A
90.00 - 94.99	A-
85.00 - 89.99	B+
80.00 - 84.99	B
75.00 - 79.99	B-
60.00 - 74.99	C
< 60.00	F

Assignments

- **Homework Assignments (25%):** You will be assigned 6 individual homework assignments over the course of the semester. These will typically involve writing some code and "knitting" a file to pdf or html in order to turn it in. Your lowest score will be dropped.

- **Exams (35%):** We will have one midterm exam worth 25% of your grade. This exam will involve writing code to complete various machine learning tasks. You can expect to be given a dataset and a series of questions to answer using the skills developed during the course. You will have three hours to complete this exam. Your final exam will be much shorter and largely conceptual in nature. You will have 90 minutes to complete the final. Exams are open everything (book, notes, internet), EXCEPT communication with others.

- **Machine Learning Models and Presentation (40%):** Over the course of the semester you will be working in a group tasked with creating several models that classify and/or predict some feature of our wine dataset. For each modeling assignment, I will provide the training and test data used to develop and measure your model's performance. Your group will have three opportunities to build a machine learning model. At each of these opportunities, the relative performance of your model determines your grade—the top group earns 20 points, 19 for second place 18 points for third place, 17 points for fourth, and 16 points for 5th. During our last class you will present your model results as though you are speaking to the managers of a large winery. Details and expectations will be clarified in class.

Course Policies

Detailed policies for this course are below. Basically, don't cheat and try to learn stuff.

- Video must be turned on in the Zoom call and names displayed.
- No private chat
- No late assignments except in **very rare** cases of personal or family emergency.
- Students with disabilities who require accommodation should notify me of the nature of accommodation you require in the first week of class. Additional support is available from the Willamette University Accessible Education Services Office, telephone 503-370-6471.
- Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee's responsibility to get all missing notes or materials.
- Every student is expected at all times to abide by the Willamette University Atkinson Graduate School of Management Honor Code.
- You must also abide by the Application to Academic Honesty as detailed in the current student handbook.

Course Schedule

Week 01, 01/11: Machine Learning Overview

- Read the syllabus
- Read Chapter 1 of this book
- Dust off your tidyverse skills

Week 02, 01/18: No Class: MLK

Week 03, 01/25: Feature Engineering I & Variable Selection

- Read the Preface and Ch. 1 of this book
- *Homework 1 due*

—*Supervised Learning*—

Week 04, 02/01: K-Nearest Neighbors

- *Homework 2 due*

Week 05, 02/08: Naive Bayes

- *Homework 3 due*

Week 06, 02/15: Logistic Regression

- *Homework 4 due*

Week 07, 02/22: Decision Trees

- *Homework 5 due*

Week 08, 03/01: Boosting and Bagging

Week 09, 03/08: Review

- *First group model due*

Week 10, 03/15: Mid-term Exam

Week 11, 03/22: No Class: Spring Break

—*Unsupervised Learning*—

Week 12, 03/29: Feature Engineering II & Dimensionality Reduction

Week 13, 04/05: Clustering

- *Second group model due*

Week 14, 04/12: The Tidy Models Package

Week 15, 04/19: Deep Learning and Ensembles

- *Final group model due*

Week 16, 04/26: Final Exam and Team Presentations