GSMDS 5004 (Spring 2020)

Applied Machine Learning

Saturdays, 9:00 to 12:00 pm

Instructor: **Jameson Watts, Ph.D.**

Email: jwatts@willamette.edu
Office Hours: Before and after class

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 pedagogies, 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 Materials:

1. Base R, Install from here
2. Latest Version of RStudio, Install from here
3. Anaconda, Install from here
4. DataCamp Classroom (DCC), Join here
5. Various free resources (links are in the schedule on the syllabus)
6. (Optional) Data Science from Scratch, Purchase here

Course Learning Objectives:

At the completion of this course, students will be able to:

1. Describe the major ethical issues facing data scientists
2. Creatively engineer new features to help with model performance
3. Implement and diagnose some of the most common supervised and unsupervised machine learning algorithms
4. Use an application interface to run a deep learning model in the cloud
5. Communicate model results in writing and in person, in a format appropriate for consumption by non-data scientists

Course Outline:

Class	Date	Class Topics	Reading and Assignments
1	01/18	Course Overview, Data Ethics & Review of Multiple Regression	Read the syllabus; Read Chapter 1 of this book
Supervised Learning			
2	01/25	Feature Engineering I & Variable Selection	Read the Preface and Ch. 1 of this book
3	02/01	k-Nearest Neighbors	DCC: k-Nearest Neighbors
4	02/08	Naive Bayes	DCC: Naive Bayes
-	02/15	**No Class**	
5	02/22	Logistic Regression	DCC: Logistic Regression
-	02/29	**No Class**	
6	03/07	Decision Trees	DCC: Classification Trees; **First group model due**
7	03/14	Mid-term Exam	
Unsupervised Learning			
8	03/21	k-Means & Hierarchical Clustering	DCC: Unsupervised Learning in R; DCC: Hierarchical Clustering
9	03/28	Feature Engineering II & Principle Component Analysis	DCC: Dimensionality Reduction with PCA; DCC: Putting it all Together with a Case Study
10	04/04	Deep Learning I	DCC: Introducing TensorFlow in R; DCC: Linear Regression using Two TensorFlow APIs; **Second group model due**
11	04/11	Deep Learning II	DCC: Deep Learning in TensorFlow: Creating a Deep Neural Network; DCC: Deep Learning in TensorFlow: Increasing Model Accuracy
12	04/18	Boosting and Ensemble Methods	Read this Vignette & this Vignette
13	04/25	Visualizing and Reporting your Models	Read this Tutorial
14	05/02	Final Exam & Team Presentations	**Final group model due**

Summary:

Assignment	Percentage
Datacamp Assignments	25%
Midterm Exam	25%
Model Performance (x3)	30%
Final Exam	10%
Group Presentations	10%
Total	100%

Grade Distribution:

>= 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%):** Homework assignments consist of completing the assigned chapters within the DataCamp Classroom and doing the assigned reading. Simply completing the assignments on time will give you full credit.

- **Exams (35%):** We will have one midterm exam worth 25% of your grade. This exam will involve writing code to complete various supervised 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.

- **Classification Models and Presentation (40%):** Over the course of the semester you will be working in a group tasked with creating a model that classifies wine. Early in the semester I will provide a training and test set upon which to measure your model's performance. Your group will have three opportunities to build and improve their 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 and 18 points for third place. During our last class you will present your model. Details and expectations will be clarified in class.

Course Policies:

- Name tents must be used for the first month of classes—I am a chronically absent-minded professor, and pretty bad with names. I promise an attempt to memorize everyone's moniker, but name tents guarantee that my deficiency does not lead to continuous embarrassment.

- Collaboration is encouraged both during and outside of the classroom. Students may work together to prepare for exams; however, each student will be on their own when taking the exams in class.

- No late assignments will be accepted under any circumstance except in 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 (www.willamette.edu/dept/disability), 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 (http://www.willamette.edu/mba/about/honorcode).

- You must also abide by the Application to Academic Honesty as detailed in the current student handbook (http://www.willamette.edu/mba/students/student-handbook).