COMP/MATH 350: Applied Machine

Learning, Fall 2018

*This syllabus is subject to change based on specific class needs, especially the schedule. Significant*

*deviations will be discussed in class.*

 Logistics

 Content

 Description

 Topics

 Learning Objectives

 Sources

 Policies

 Assessment

 Assignments

 Grading

 Workload

 Schedule

 Monmouth College Services

Logistics

 Class meetings: MTWF 2:00 PM - 2:50 PM in CSB 303

 Instructor: Robert Utterback

 Office: CSB 342

 Phone: 309.457.2202

 Website: https://robertutterback.github.io

 Email: rutterback@monmouthcollege.edu

 Office hours: M 9-10 AM. Tu 3-4 PM. W 3-4 PM. Th 9-11 AM. F 9-10 AM. By appointment (see the

schedule on my webpage)

 Website: https://robertutterback.github.io/courses/comp350ml/f18/

 Credits: 1 course credit

 Prerequisites: COMP 160 and MATH 260.

 Feedback: At any time during the course you can use this to provide anonymous suggestions, criticism,

praise, etc.

Content

Description

An introduction to the hot topics of machine learning, data science and data mining. The course aims to supply

students with a useful toolbox of machine learning techniques that can be applied to real-life data. Techniques

may include logistic and linear regression, SVMs, decision trees, neural networks, and clustering. The focus will

/

be on developing important skills in preparing data and selecting and evaluating models, though we will delve

into the mathematical intuition behind each model.

Topics

Possible topics include:

 Python for data science and visualization

 Linear models for regression

 Linear models for classification

 Preprocessing, feature selection and engineering

 Support vector machines

 Trees, forests, and ensembles

 Gradient boosting

 Model Evaluation and Selection

 Dimensionality reduction

 Clustering

 Word embeddings

 Neural Networks

Learning Objectives

 An understanding of the types of machine learning and the intuition behind the techniques listed above

 An ability to analyze real-world data and prepare it for learning/mining via data cleaning

 An ability to add, remove, and modify data features and evaluate their effect on ML models

 An ability to compare and evaluate the performance of several ML models on a particular dataset

 An ability to communicate algorithmic findings and effectively present results through visualization

Sources

The required course textbook is:

Kuhn, Max and Johnson, Kjell. *Applied Predictive Modeling*. Springer. 2013. ISBN-13: 978-1-4614-6848-6.

I also recommend, but do not require:

Guido, Sarah and Muller, Andreas. *Introduction to Machine Learning with Python*. O’Reilly. 2016. ISBN-13: 978-

1449369415.

Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron. *Deep Learning*. MIT Press. 2016. ISBN:

9780262035613.

Policies

 **Late assignments**: You have each been allotted a *total* of *5* late days. You may apply these to any

problem sets you see fit and turn in your solutions with no penalty. However, you may use at most 2 on any

individual assignment. The whole point here is to give you some flexibility that allows for things like

illnesses, long trips, and the like. I am unlikely to grant further extensions.

 **Academic dishonesty**: Monmouth College’s official policy on academic dishonesty can be found here.

**You** are responsible for reading and complying with that policy.

*In this course, any violation of the academic honesty policy will have varying consequences depending on*  /

*the severity of the infraction as judged by the instructor. Minimally, a violation will result in an “F” or 0 points on the assignment in question. Additionally, the student’s course grade may be lowered by one letter grade. In severe cases, the student will be assigned a course grade of “F” and dismissed from the class. All cases of academic dishonesty will be reported to the Associate Dean who may decide to recommend further action to the Admissions and Academic Status Committee, including suspension or dismissal. It is assumed that students will educate themselves regarding what is considered to be academic dishonesty, so excuses or claims of ignorance will not mitigate the consequences of any violations*

 **Collaboration**: We encourage you to make use of the resources available to you – it is fine to seek help from a friend, tutor, instructor, internet, etc. However, *copying of answers and any act worth of the label* *“cheating” is never permissible*! In addition to listing your sources and collaborators, you should be producing your own writeup in your own words. By “your own words,” we mean you should be producing the text yourself, without some external aid. Verbatim copying of text is specifically disallowed, but so is taking a source and rearranging some phrases and changing some variable names to create a derivative version! Such behavior is definitely NOT “using your own words.” It does not matter if you helped contribute to this source text with others, since then you are still not the sole author of the text. The point of collaborating on an assignment is not to produce a jointly authored set of solutions, since that violates the course policies. Instead, it is to help you solve the problems, which sometimes involve a bit of creativity.

After you have jointly come up with the ideas you need to solve the problems, though, you should part ways with your group and sit down to do the writing by yourself. I also advise against sharing the writeup you submit with others, since if someone else uses your text as a source for their own solution (with or without your permission), you will also be implicated in the violation of the academic integrity policy. In any case, if two nearly identical solutions are received, we have no way of tell which is the original, and the policy is to not award credit for either submission.

 **Electronic devices**: Do not use your phone in class. Keep it on silent or leave it at home. Any computer or tablet usage should be related to the course. Other usage is rude and distracting to others.

 **General expectations**: In short, I expect you to be respectful of others and take responsibility for your own learning. You are here to learn, so work hard and be professional.

 Just attending class is not sufficient to truly learn the material. Read the text, use the resources available at Monmouth College, and go beyond the material.

 If you miss class, you are responsible for everything covered on that day. College is, in some sense, your job. Take pride in creating quality work. Staple your assignments, label problems, and present your answers neatly and orderly.

 Your job is to convince me that you have learned the material – show your work! Even if you do not know a particular answer, guide me through your thought process.

Assessment

Assignments

The course workload is as follows:

|  |  |  |
| --- | --- | --- |
| **Category** | **Number of Assignments** | **Final Grade Weight** |
| Homework | 5–7 | 50% |

/

|  |  |  |
| --- | --- | --- |
| **Category** | **Number of Assignments** | **Final Grade Weight** |
| Midterm | 1 | 20% |
| Final | 1 | 20% |
| Participation | - | 10% |

Most (probably all) homework assignments will involve programming. Each exam focuses primarily, but not necessarily exclusively, on material covered since the previous exam. In other words, the final exam may include one or two questions from first-half material.

Your participation grade is based on a variety of activities. During class I will often make sure of the Socrative app, so you’ll need to install this on your phones. Participating in Socrative questions and with in-class group activities is required for a decent participation grade; a full grade also includes asking questions either in class or in office hours.

Grading

Your final grade is based on a weighted average of particular assignment categories, with weights shown above. You can estimate your current grade based on your scores and these weights. You may always visit the instructor *outside of class* to discuss your current standing.

This courses uses a standard grading scale. Assignments and final grades will not be curved except in rare cases when its deemed necessary by the instructor. Percentage grades translate to letter grades as follows:

|  |  |
| --- | --- |
| **Score** | **Grade** |
| 94–100 | A |
| 90–93 | A- |
| 88–89 | B+ |
| 82–87 | B |
| 80–81 | B- |
| 78–79 | C+ |
| 72–77 | C |
| 70–71 | C- |
| 68–69 | D+ |
| 62–67 | D |
| 60–61 | D- |
| 0–59 | F |

You are always welcome to challenge a grade that you feel is unfair or calculated incorrectly. Mistakes made in your favor will never be corrected to lower your grade. Mistakes made not in your favor will be corrected. *Basically, after the initial grading your score can only go up as the result of a challenge*.

Workload

/

The weekly workload for this course will vary by student and over the semester, but on average should be about 12 hours per week. The follow table provides a rough estimate of the distribution of this time over different course components for a 16 week semester.

|  |  |  |
| --- | --- | --- |
| **Category** | **Total Time** | **Time/Week (Hours)** |
| Lectures | 55 | 3.5 |
| Homework | 72 | 4.5 |
| Exam Study | 27 | 1.5 |
| Reading+Unstructured Study |  | 2.5 |

12

Schedule

The following **tentative** calendar should give you a feel for how work is distributed throughout the semester. Assignments and events are listed in the week they are due or when they occur. *This calendar is subject to change based on the circumstances of the course*.

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Topic** | **Assignment** |  |
| Wed 08/22 | Intro, What is Machine Learning |  |  |
| (Week 1) | (pdf) |
| Fri 08/24 | ML Principles; Python | Do a Python tutorial, e.g. this one |  |
| Mon 08/27 | Essential Python Libraries | HWK 1 out |  |
| (Week 2) |
| Tue 08/28 | Python Visualization Basics | Read APM 1-2 |  |
| Wed 08/29 | Classification and Regression | Read IMLP 1 |  |
| Case Studies |
| Fri 08/31 | Model Complexity and Sources of | Read APM 4.1-4.2 |  |
| Error |
| Mon 09/03 | Model Tuning | Read APM 4.3-4.4 |  |
| (Week 3) |
| Tue 09/04 | Comparing Models | Read 5.1-5.2 |  |
| Wed 09/05 | Linear and Ridge Regression | HWK 1 due (Solutions), HWK 2 out, |  |
| Read APM 6.1-6.2 |
| Fri 09/07 | Understanding Regularization | Read APM 6.4 |  |
| Mon 09/10 | Linear Models for (Binary) | Read APM 12.1-12.2 |  |
| (Week 4) | Classification |
| Tue 09/11 | Multiclass Classification | Read 12.5 |  |
| Wed 09/12 | Computational Considerations; |  |  |
| SGD |
| Tue 12/11 3:00 | Final Exam |  | / |
| PM |

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Topic** | **Assignment** |  |
| Fri 09/14 | Scaling Data | Read APM Ch 3 |  |
| Mon 09/17 | Review and HWK 2 Questions | HWK 3 out |  |
| (Week 5) |
| Tue 09/18 | Processing Pipelines and Feature |  |  |
| Engineering |
| Wed 09/19 | Feature Engineering | HWK 2 due (Solutions), |  |
| Fri 09/21 | Imputation | Read APM 3.4 |  |
| Mon 09/24 | Feature Selection | Read APM 19 |  |
| (Week 6) |
| Tue 09/25 | Support Vector Machines (no | Read APM 7.3 |  |
| slides) |
| Wed 09/26 | SVM Kernels | Read APM 13.4 |  |
| Fri 09/28 | Trees and Forests | Read APM 8-1-8.2, 14.1 |  |
| Mon 10/01 | Homework 2 Review and Random | Read 8.4-8.5, 14.3-14.4 |  |
| (Week 7) | Forests |
| Tue 10/02 | Ensembles and Gradient Boosting | Read 8.6, 14.5 |  |
| Wed 10/03 | Stacking |  |  |
| Fri 10/05 | Calibration | HWK 3 checkpoint due, Practice |  |
| Midterm out |
| Mon 10/08 | Review | Practice Midterm Solutions |  |
| (Week 8) |
| Tue 10/09 | Review |  |  |
| Wed 10/10 | Midterm (Solutions) |  |  |
| (10/12–10/15) | (Fall Break) |  |  |
| Tue 10/16 | Midterm Review | HWK 3 due (Solutions) |  |
| (Week 9) |
| Wed 10/17 | Mean, Median, Mode |  |  |
| Fri 10/19 | Model Evaluation Metrics | Read APM 11 & 5 |  |
| Mon 10/22 | Imbalanced Data | Read APM 16 |  |
| (Week 10) |
| Tue 10/23 | Synthetic Data Generation |  |  |
| (Wed 10/24) | (Mentoring Day – No class) |  |  |
| Fri 10/26 | Dimensionality Reduction |  |  |
| Mon 10/29 | (Class Cancelled) |  |  |
| (Week 11) |
| Tue 12/11 3:00 | Final Exam |  | / |
| PM |

|  |  |  |
| --- | --- | --- |
| **Date** | **Topic** | **Assignment** |
| Tue 10/30 | Dimensionality Reduction | Read IMLP 142-157, 165-170, HWK 4 |
| (continued) | out |
| Wed 10/31 | (Class Cancelled) |  |
| Fri 11/02 | Clustering | Read IMLP 170-193 |
| Mon 11/05 | (Class Cancelled) |  |
| (Week 12) |
| Tue 11/06 | Mixture Models and Clustering | Read IMLP 193-211 |
| Evaluation |
| Wed 11/07 | Unsupervised Clustering |  |
| Evaluation |
| Fri 11/09 | More Clustering Evaluation |  |
| Mon 11/12 | NMF and Outlier Detection | Read IMLP 158-170 |
| (Week 13) |
| Tue 11/13 | Working with text data | Read IMLP 325-336 |
| Wed 11/14 |  | Read IMLP 337-349 |
| Fri 11/16 | LSA | HWK 4 due, Read IMLP 349-358 |
| Mon 11/19 | Topic Models | HWK 5 out |
| (Week 14) |
| Tue 11/20 | Word and document embeddings |  |
| (11/21–11/25) | (Thanksgiving Break) |  |
| Mon 11/26 | (Class Cancelled) |  |
| (Week 15) |
| Tue 11/27 | Neural networks | Skim DL Ch. 6 |
| Wed 11/28 |  | Read APM 7.1 and 13.2 |
| Fri 11/30 | NN in Practice | HWK 6 out |
| Mon 12/03 | Advanced NN | Watch these, Practice Final |
| (Week 16) |

Tue 12/04

|  |  |  |
| --- | --- | --- |
| Wed 12/05 | Review | Practice Final Solutions |
| Tue 12/11 3:00 | Final Exam |  |
| PM |

Monmouth College Services

 The Teaching and Learning Center offers FREE resources to assist Monmouth College students with their academic success. Programs include supplemental instruction for difficult classes, drop-in and appointment tutoring, and individual academic coaching. The TLC is here to help students excel

/

academically. TLC services are not just for struggling students, but can assist all students to get better grades, practice stronger study skills, and manage time.

Make an appointment with Kam Williams, Director of Academic Support Programs and Student Disability Services, at the TLC on the 2nd floor of Poling Hall. The department phone number is 457-2257, or contact the department online at http://ou.monmouthcollege.edu/academics/teaching-learning-center/. They can also be reached via email at: tlc@monmouthcollege.edu

 Disability Support Services: If you have a disability or had academic accommodations in high school or another college, you may be eligible for academic accommodations at Monmouth College under the Americans with Disabilities Act (ADA). Monmouth College is committed to equal educational access.

Students can meet with Kam Williams about accommodations at the Teaching and Learning Center (TLC). The TLC is located on the 2nd floor of Poling Hall. For more information, call 309-457-2257 or connect online at http://ou.monmouthcollege.edu/life/disability-services/default.aspx.

/