MSSC 6250 Statistical Machine Learning

Instructor: Dr. Cheng-Han Yu

Spring 2024

E-mail: cheng-han.yu@marquette.edu
Office Hours: TuTh 4:50 - 5:50 PM, Wed 12 - 1 PM or by appointment
Office: Cudahy Hall 353
Teaching Assistant (TA): No TA :(
Web: d2l.mu.edu/d2l/home
Class Hours: TuTh 3:30 - 4:45 PM
Class Room: Cudahy Hall 143

Course Objectives

The course covers supervised learning and unsupervised learning models and algorithms. Supervised learning methods include various regression and classification methods, and unsupervised learning methods involves dimension reduction and clustering techniques. Topics include Bayesian linear regression, shrinkage and regularization, regression splines, Gaussian processes, logistic regression, discriminant analysis, nearest neighbors, tree-based methods, principal components, K-means, Gaussian mixture clustering, neural networks, etc.

Prerequisites

- MATH 4720 (Intro to Statistics), MATH 3100 (Linear Algebra) and MATH 4780 (Regression Analysis)
- Having taken MATH 4700 (Probability) and MATH 4710 (Statistical Inference) or more advanced ones is strongly recommended.
- This course is supposed to be taken in the *last* semester for the applied statistics (APST) master students. Talk to me if you are not sure whether or not this is the right course for you.

Required Textbook

- (**ISL**) *An Introduction to Statistical Learning, 2nd edition*, by James et al. Publisher: Springer. (Undergraduate to master level, R and Python code)

Optional References

- (**PML**) *Probabilistic Machine Learning: An Introduction*, by Kevin Murphy. Publisher: MIT Press. (Master to PhD level, lots of mathematics foundations, Python code)
- (**PMLA**) *Probabilistic Machine Learning: Advanced Topics*, by Kevin Murphy. Publisher: MIT Press. (PhD level, more probabilistic-based or Bayesian)
- (**ESL**) *The Elements of Statistical Learning, 2nd edition*, by Hastie et. al. Publisher: Springer. (PhD level, more frequentist-based)

Learning Management System (D2L)

- Any new announcement will be posted in **News** in **Course Home**.
- Please start with **Content** > **Course Information** that includes the syllabus, homework schedule and technology issue solutions.
- Please carefully follow Week modules (**Content** > **Week 1 to Week 16**) that includes course materials and homework.
- You can go to **Content** > **Start Here: Course Information** or https://www.marquette.edu/remote-learning/d2l.php to learn more about D2L.

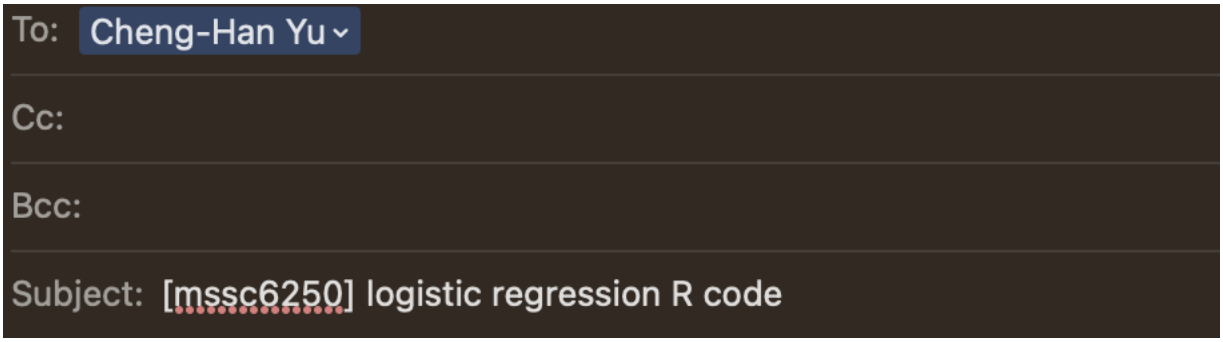

Figure 1: Email Subject Line Example

Office Hours

- My in-person office hours are TuTh 4:50 - 5:50 PM, Wed 12 - 1 PM in Cudahy Hall room 353.
- We can schedule an online meeting via Microsoft Teams if you need/like.

E-mail Policy

- I will attempt to reply your email quickly, at least **within 24 hours**.
- **Expect a reply on Monday if you send a question during weekends**. If you do not receive a response from me within two days, re-send your question/comment in case there was a "mix-up" with email communication (Hope this won't happen!).
- Please start your subject line with **[mssc6250]** followed by a clear description of your question. See an example in Figure 1.
- Email etiquette is important. Please read this article to learn more about email etiquette.
- I am more than happy to answer your questions about this course or statistics in general. However, due to time constraint, I may choose **NOT** to respond to students' e-mail if
  1. The student could answer his/her own inquiry by reading the syllabus or information on D2L.
  2. The student is asking for an extra credit opportunity. The answer is "no".
  3. The student is requesting an extension on homework. The answer is "no".
  4. The student is asking for a grade to be raised for no legitimate reason. The answer is "no".
  5. The student is sending an email with no etiquette.

Grading Policy

- Your final grade is earned out of **1000 total points** distributed as follows:
  - **Homework: 500 pts**
  - **In-class Activity: 200 pts**
  - **Final project presentation and/or written report: 300 pts**
- You will **NOT** be allowed any extra credit projects/homework/exam to compensate for a poor average. Everyone must be given the same opportunity to do well in this class. Individual exam will **NOT** be curved.
- The final grade is based on your percentage of points earned out of 1000 points and the grade-percentage conversion Table. $[x, y)$ means greater than or equal to x and less than y. For example, 94.1 is in $[93, 100]$ and the grade is A and 92.8 is in $[90, 94)$ and the grade is A-.

Table 1: Grade-Percentage Conversion

Grade	Percentage
A	[94, 100]
A-	[90, 94)
B+	[87, 90)
B	[83, 87)
B-	[80, 83)
C+	[77, 80)
C	[70, 77)
F	[0, 70)

- You may use any programming language to do your homework and/or your project.

Homework

- The homework assignments are *individual*. You should submit your own work.
- Homework will be assigned through the course website.
- To submit your homework, please go to **D2L > Assessments > Dropbox** and upload your homework in **PDF** format.
- You will have *at least one week* to complete your assignment.
- **No late or make-up homework**.

In-Class Activity

- There will be 3 to 4 in-class activities.
- Students will learn from each other by presenting and discussing the assigned topics.
- More details about the in-class activities will be released later.

Project

- The final project includes two parts: written report and oral presentation.
- **You have to participate (in-person) in the final presentation in order to pass the course**.
- More details about the written report and oral presentation will be released later.

Academic Integrity

- This course expects all students to follow University and College statements on academic integrity.

- **Honor Pledge and Honor Code**: *I recognize the importance of personal integrity in all aspects of life and work. I commit myself to truthfulness, honor, and responsibility, by which I earn the respect of others. I support the development of good character, and commit myself to uphold the highest standards of academic integrity as an important aspect of personal integrity. My commitment obliges me to conduct myself according to the Marquette University Honor Code.*

Accommodation

If you need to request accommodations, or modify existing accommodations that address disability-related needs, please contact Disability Service.

Attendance and COVID-19

- It is your responsibility as a Marquette University student to protect the health and safety of our community in this course. The University has mandated vaccination for all students for the academic year 2022-2023. COVID Cheq will still be used, but those fully vaccinated that provide documentation will no longer receive the daily symptom screening survey. Instead, they will receive a daily email providing a green check. You may be required to show your automated green check to show clearance for entry into campus buildings. If you are experiencing symptoms related to COVID-19, you must follow the link in the automated email to report symptoms.
- Visit What to do if you are exposed to COVID-19 or test positive website for university guidelines on the best course of action.
- Visit guidance on Spring 2024 Class attendance, withdrawal, and grading
  - Students are responsible for contacting instructors prior to the missed class session to indicate absence and the need to make up classwork/assignments.
  - Students requesting make up classwork/assignments are required to provide the COVID Cheq "stop sign" to confirm inability to attend class.

Tentative Course Schedule

Week 1, 1/15 - 1/21: Overview of Statistical Learning, Bias-variance trade-off

Week 2, 1/22 - 1/28: Linear Regression

- **Drop deadline 1/24 11:59 PM**

Week 3, 1/29 - 2/4: Ridge Regression and Cross-Validation

Week 4, 2/5 - 2/11: Feature Selection and Lasso

Week 5, 2/12 - 2/18: Splines, Generalized Additive Models

Week 6, 2/19 - 2/25: Bayesian Linear Regression

Week 7, 2/26 - 3/3: Generalized Linear Models: Logistic Regression and Poisson Regression

Week 8, 3/4 - 3/10: Generative Models: Discriminant Analysis, Naive Bayes

Week 9, 3/11 - 3/17:

- **No class this week (Spring Break)**

Week 10, 3/18 - 3/24: K-Nearest Neighbors Regression and Classification

Week 11, 3/25 - 3/31: Gaussian Process Regression and Classification

- **No class on 3/28 (Easter Break)**

Week 12, 4/1 - 4/7: Support Vector Machines

Week 13, 4/8 - 4/14: Tree Methods: CART, Bagging, Random Forests, Boosting

- **Withdrawal Deadline 4/12**

Week 14, 4/15 - 4/21: Principal Component Analysis and Recommender Systems

Week 15, 4/22 - 4/28: K-Means Clustering, Model-based Clustering

Week 16, 4/29 - 5/5: Deep Learning and Neural Networks

Week 17, 5/6 - 5/12:

- **Project Presentation: Thursday, 5/9, 8 - 10 AM**
- **Final grade submission 5/14 by noon**

*** I reserve the right to make changes to the syllabus.**