# CS 5350/6350: Machine Learning

| | | | |
|---|---|---|---|
| **Instructor:** | Shandian Zhe | **Time:** | Tue & Thu 09:10 - 10:30am |
| **Email:** | zhe@cs.utah.edu | **Place:** | MEK 3550 |

**Course Page:**

- https://www.cs.utah.edu/~zhe/teach/cs6350.html

**Teaching Mentee and Assistant:**

- Da Long (u1368737@utah.edu)
- Tushar Gautam (tushar.gautam@utah.edu)
- Caleb Johnson (calebdeejohnson@gmail.com)

**Office Hours:**

- Instructor: Tue & Thu, 12:30pm - 1:50 pm, MEB 3466
- Da Long: Mon 12pm - 2pm & Fri 8am-9am Zoom
- Tushar Gautam: Fri 10am-1pm Zoom
- Caleb Johnson: Wed 11am - 1pm & Fri 9am-10am Zoom

**Description:**

This course introduces basic knowledge of machine learning. Topics consist of several fundamental, and widely successful supervised/unsupervised learning algorithms, such as decision trees, perceptrons, (deep) neural networks, kernel methods, support vector machines and probabilistic methods. After taking this class, we hope that you will (1) *understand machine learning ideas and paradigms*, (2) *be able to identify appropriate machine learning problems for your research or applications*, and (3) *be able to design machine learning models and to implement learning algorithms (with appropriate tools if necessary)*.

**Books:** There are no designated textbooks for this course. The lecture slides will cover all the content. Here we list several books to further extend the depth and breadth of the topics we will discuss in the class.

- Kevin Patrick Murphy, Machine Learning: a Probabilistic Perspective. MIT Press, 2012.
- Christopher Bishop, Pattern Recognition and Machine Learning. Springer 2007.
- Hal Daumé, A Course in Machine Learning.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning. Springer 2008.
- David J.C. MacKay, Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.

**Prerequisites:** "C-" or better in (CS 3500 AND (MATH 2270 OR MATH 2250) AND (CS 3130 OR ECE 3530)) AND (Full Major status in Computer Science OR Computer Engineering). Corequisites: "C-" or better in (CS 4150 OR CS 3100). Basically, we assume that you

- know basics of probability theory and statistics,
- are familiar with linear algebra, vector/matrix derivatives,
- have algorithmic design, programming and debugging skills.

**Getting Help:** Take advantage of the instructor and TA office hours (posted on the course web page). We will work hard to be accessible to students. Please send us emails if you need to meet outside of office hours. Don't be shy if you don't understand something: come to office hours, send email, or speak up in class! Students are encouraged to use a discussion group for additional questions outside of class and office hours. The class will use **Canvas Q/A**. Feel free to post questions regarding any questions related to class: homeworks, schedule, material covered in class. Also feel free to answer questions, the instructors and TAs will actively answer questions. However, **do not post potential answers for homework assignments**. Such posts will be immediately removed, and not be answered. All important announcements will be made through the discussion group, there is otherwise no class mailing list.

**Tentative Course Outline:**

- Supervised machine learning
  - Decision trees
  - Overfitting and cross-validation
  - Boosting and ensemble learning
  - Linear regression: least mean square (LMS) approach
  - Linear classifier
    - Perceptron
    - Support vector machines (SVM)
    - Probabilistic learning
      - Naive Bayes classifier
      - Logistic regression
  - Nonlinear classifier
    - Nearest neighbour classifier
    - Kernel tricks
      - Kernel perceptron
      - nonlinear SVM
    - Deep neural networks
- Unsupervised learning
  - K-means clustering and its variants
  - Gaussian mixture models
- Practical guidance: Empirical suggestions, Tensorflow/PyTorch tutorials

**Grading Policy:**

- Homeworks (50%): 6 homeworks. The homeworks will consist of both analytical problems and programming assignments. The programming assigments usually require you to implement particular machine learning models and test them in real-world/sythetic datasets. Each homework might have a different number of points, depending on the work amount. You are highly encouraged to use

Python, MATLAB and/or R for the programming portion of the assignments or projects. However, you can choose any other programming language. But you should guarantee that your programs can be compiled and run on the CADE machines; otherwise you will NOT receive credits from the programming parts. In the programming assignments, feel free to use existent libraries (e.g., Numpy and Scipy in Python) to finish matrix computation and optimization (unless the homework says it is not allowed). *Your are never allowed to call APIs (e.g., scikit-learn) that directly fulfills the required models/algorithms.*

- Course project (30%): One project that uses machine learning techniques to address practical problems. A Github repository is required to maintain and update the project. **Please set your Github repository to private and only allow the TM/TAs to access.** The mid-term and final reports are required to submit. For details, please refer to the course web page `https://www.cs.utah.edu/~zhe/teach/cs6350-proj`
- Final exam (20%): It is open-book. You can bring whatever materials that you think will help the exam, but no computers, cell phones, or tablets are allowed.

**Letter Grade Mapping:** We do not curve your grade, but the letter grade mapping varies a bit from undergraduate level (CS 5350) to graduate level (CS 6350).

Table 1: CS5350

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 88-100 | A | 71-75 | B | 56-60 | C | 41-46 | D |
| 82-87 | A- | 66-70 | B- | 51-55 | C- | 36-40 | D- |
| 76-81 | B+ | 61-65 | C+ | 46-51 | D+ | 0-35 | E |

Table 2: CS6350

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 91-100 | A | 74-78 | B | 59-63 | C | 44-49 | D |
| 85-90 | A- | 69-73 | B- | 54-58 | C- | 39-43 | D- |
| 79-84 | B+ | 64-68 | C+ | 49-54 | D+ | 0-38 | E |

We will round up your score to the closest integer values. For example, 90.5-90.9 will be treated as 91, and 90.1-90.4 will be treated as 90.

**Important Dates:**

Final Exam ..... Thursday, December 15, 2022 8:00 – 10:00 am, MEK 3550

**Late Policy:** Each assignment must be turned in through **Canvas** by the designated deadline, usually 11:59pm, to receive the full credit. If the deadline is missed, the late submission will have 10% penalty. Then late submissions in every subsequent 24 hours will lose another 10% credit. For example, an 10 points assignment will have 2 points penalty, if it is submitted 30 hours late. However, if the assignment is not turned in until the other assignment have been graded and returned or 48 hours after the deadline, 0 grade will be given.

Assignments will be posted far enough ahead of time that I will not be able to make exceptions if a student falls ill. The exception will be prolonged illness accompanied by a doctor's note.

If you believe there is an error in grading homeworks, **you may request a regrading within one week of receiving your grade**. Requests must be made by email to instructor, explaining clearly why you think your solution is correct.

**Homework Submission:** We only accept homeworks written with LaTex. It is something that everyone should know for research and writing scientific documents. This linked directory (http://www.cs.utah.edu/~jeffp/teaching/latex/) contains a sample .tex file, as well as what its .pdf compiled outcome looks like. It also has a figure .pdf to show how to include figures. We will release the template text file along with each homework assignment for convenient edit. It is recommended to use online latex editor overleaf (https://www.overleaf.com/project).

**Attending the Final Exam:** If you are unable to take the final exam at the scheduled time, please contact the instructor ahead, and we might be able to schedule another time for you.

If you miss the final exam due to emergency, please notify the instructor within 3 days of the exam and provide proof document. Then we can consider scheduling a make-up exam.

**Academic Policy:** This course follows School of Computing (SoC) Polices/Guidelines (https://handbook.cs.utah.edu/). Please read it carefully. All the registered students are required to sign the Acknowledgement Form (http://www.cs.utah.edu/docs/misc/SoC_ack_form.pdf), and turn into the department office before any homework is graded.

If a student is caught cheating on a homework or test, they will receive a failing grade for the course. For a detailed description of the university policy on cheating, please see the University of Utah Student Code: https://regulations.utah.edu/academics/6-400.php

**Collaboration policy**: For assignments, students can discuss answers with anyone, but must write their own code, proofs and write-ups. If you collaborated with another student such that you expect the answers may look similar, you must explain explicitly in the homework submission to what extent you have collaborated. Students' homework submissions appearing too similar to others and without collaboration explanation will receive 0 grades.

For the final exam, discussing with anyone (except for instructors/TAs) during the exam period is NOT allowed, and will result 0 grade on the exam.

**Students with Disabilities**

Please let me know at your earliest convenience. The University of Utah seeks to provide equal access to its programs, services, and activities for people with disabilities. If you need accommodations in this class, reasonable prior notice needs to be given to the Center for Disability Services, 162 Olpin Union Building, 581-5020 (V/TDD). CDS will work with you and the instructor to make arrangements for accommodations.