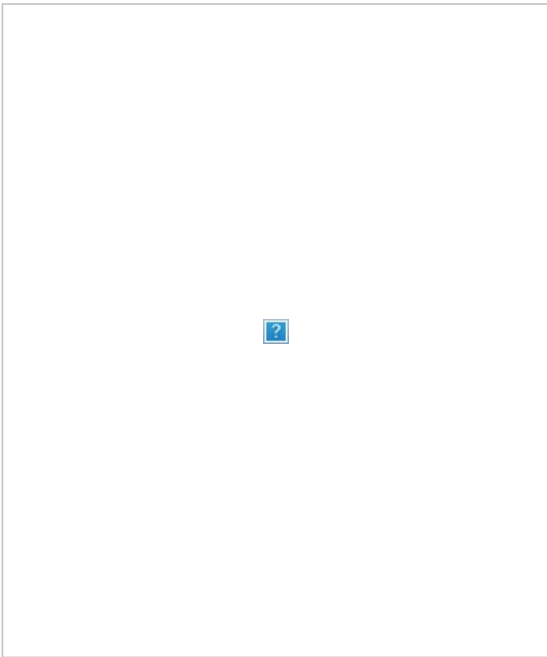# CSCI 567: Machine Learning

Basic Information

- **Lecture time:** Thursdays 5:00 pm to 7:20 pm, followed by discussion from 7:30 pm to 8:20 pm
- **Lecture place:** SGM 123 and Zoom
- **Instructor:** Vatsal Sharan (vsharan)
- **TAs:** Liyu Chen (liyuc), Ameya Godbole (ameyagod), I-Hung Hsu (ihunghsu), Sampad Mohanty (sbmohant), Bhavya Vasudeva (bvasudev), Wang (Bill) Zhu (wangzhu)
- **CPs & Graders:** Jiayi Chang (jiayicha), Abhay Iyer (adiyer), Sahana Kalakonda (skalakon), Syed Hameed Mohamed (syedhame), Dipak Wani (dpwani), Rachitha Kagalvadi Shivappa [Grader] (kagalvad)
- **Office Hours:** Link to all office hours.
- **Communication:** All inquiries which do not pertain to a specific member of the course staff should be sent via ed Discussion (see below). USC email-ids of all staff members are in parantheses above.
- **ed Discussion:** We will be using ed for all course communications (regarding homework, project, course scheduling, etc). Please feel free to ask/answer any questions about the class on ed. You can post privately on ed to contact the course staff for any reason. You should be enrolled in ed automatically.

Course Description and Objectives

Is this what we'll learn to do in this class? Or is this what we'll learn not to do?

The chief objective of this course is to introduce standard statistical machine learning methods, including but not limited to various methods for supervised and unsupervised learning problems. Particular focus is on the conceptual understanding of these methods, their applications, and hands-on experience.

Prerequisites

(1) Undergraduate level training or coursework on linear algebra, (multivariate) calculus, and basic probability and statistics;
(2) Basic skills in programming with Python;
(3) Undergraduate level training in the analysis of algorithms (e.g. runtime analysis).

Syllabus and Materials

The following is a tentative schedule. The quiz timings are fixed, but the rest of the content will likely change as the course continues. We will also post lecture notes and assignments here.

Please refer to Ed Discussion for recommended readings.

Date	Topics	Lecture notes	Homework/ Practice problems
08/25	Lecture: Introduction, Linear regression; Optimization algorithms	Lecture slides, Optimization Colab	Linear algebra

Date	Topic	Materials	Assignments
	Discussion: Linear algebra & Numpy review, Part I	Discussion slides, Linear algebra Colab I	questions I
09/01	Lecture: Linear classifiers; Perceptron; Logistic regression Discussion: Probability review	Lecture slides, Optimization Colab Discussion notes, Gaussians Python notebook	Probability questions HW1, HW1 solutions
09/08	Lecture: Generalization; Nonlinear basis; Regularization Discussion: Linear algebra & Numpy review, Part II	Lecture slides, Nonlinear functions Colab Discussion notes, Linear Algebra Colab II	Linear algebra questions II
09/15	Lecture: L1 regularization; Kernel methods Discussion: HW1 review	Lecture slides Discussion notes	HW2, HW2 solutions
09/22	Lecture: SVM Discussion: Problem discussion for Quiz 1	Lecture slides, SVM Colab Discussion notes	Quiz 1 practice Quiz 1 practice (solutions)
09/29	Lecture: Multiclass classification; Neural Networks Discussion: HW2 review	Lecture slides	
10/06	Lecture: Quiz 1 No Discussion session		HW3, HW3 solutions
10/13	Fall recess		
10/20	Lecture: Neural networks for images and sequences (and Markov models) Discussion: Quiz 1 review	Lecture slides Discussion notes	
10/27	Lecture: Decision trees; Ensemble methods Discussion: HW3 review	Lecture slides Discussion notes	HW4, HW4 solutions
11/03	Lecture: Dimensionality reduction and visualization; PCA Discussion: Project overview	Lecture slides Discussion notes	
11/10	Lecture: Clustering; k-means; Gaussian mixture models; EM Discussion: Evaluation metrics (precision, recall etc.)	Lecture slides, EM demo (exercise) Discussion notes	
11/17	Lecture: Density estimation; Generative models & Naive Bayes; Multi-armed bandits; Responsible ML Discussion: HW4 review	Lecture slides Discussion notes	
11/24	Thanksgiving		
12/01	Lecture: Quiz 2 No Discussion session		Practice problems, Practice problem solutions
12/14	Project report due		Project logistics

Requirements and Grading

1. **4 homeworks** worth 40% of the grade. The homeworks will be a combination of theoretical and exploratory programming questions. They should be done in **groups of 2**. One late day will be available to every student for the homeworks. Note that a group can submit late by using a late day of any one group member, therefore if you work with the same partner throughout your group effectively has 2 late days available (but you cannot apply both late days to one homework).
2. **Two quizzes** during class hours worth 20% each. The quizzes will test conceptual understanding of the material covered in the lectures, discussions and assignments.
3. A **course project** worth 20%. The project should be in **groups of 4 students**. It will be an applied project most likely be based on a Kaggle competition. Each group will be asked to submit a project report. More information will be released later.
4. **Contributions to the class** (Discretionary Grade Bumps): You are encouraged to help your fellow classmates when possible and improve everyone's learning experience, such as by responding to Ed Discussion questions when you know the answer. At the end of the course, we will bump up grades of those students who had the most positive impact on the class, according to the (quite subjective) judgement of the course staff.

Resources and related courses

1. There is no required textbook for this class, but the following books are good supplemental reading for many parts.
   - Probabilistic Machine Learning: An Introduction [PML] by Kevin Murphy. Available online here.
   - Elements of Statistical Learning [ESL] by Trevor Hastie, Robert Tibshirani and Jerome Friedman. Available online here.
   - PATTERNS, PREDICTIONS, AND ACTIONS A story about machine learning, by Moritz Hardt and Benjamin Recht. Available online here.
   - (for more of the theory) Understanding Machine Learning: From Theory to Algorithms, by Shai Shalev-Shwartz and Shai Ben-David. Available online here.
2. This course draws heavily from several other related courses, particular the previous iteration of this class by Prof. Haipeng Luo and other USC faculty:
   - Haipeng Luo's class at USC. [website]
   - CS229 at Stanford. [website]

- Greg Valiant's class at Stanford. [website]

Helpful reminders

Collaboration policy and academic integrity: Our goal is to maintain an optimal learning environment. You can discuss the homework problems at a high level with other groups, but you should not look at any other group's solutions. Trying to find solutions online or from any other sources for any homework or project is prohibited, will result in zero grade and will be reported. To prevent any future plagiarism, uploading any material from the course (your solutions, quizzes etc.) on the internet is prohibited, and any violations will also be reported. Please be considerate, and help us help everyone get the best out of this course.

Please remember the Student Conduct Code (Section 11.00 of the USC Student Guidebook). General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty.

Students with disabilities: Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to the instructor as early in the semester as possible.