CSCI-567: Machine Learning

Summer 2022

Prof. V. Adamchik

Course Description:

Machine learning (ML) is a set of algorithms that allow machines to learn (the way humans do) from experience, by extracting useful information and taking the decision based upon the data analysis. The essence of ML is ability to learn from data, identify patterns and build successful predictive models for the unknown datasets. Over the past two decades, machine learning has become increasingly central both in artificial intelligence (AI) as an academic field, and in the technology industry. ML is generally considered separate from AI which is more about building systems to do intelligent things. Machine learning is a specific subset of AI which primarily deals with creation of algorithms which learn with experience and improve themselves over time by feeding on data.

This course provides students with an in-depth introduction to the theory and practical algorithms for machine learning from a variety of perspectives. It covers some of the main models and algorithms for regression, classification, clustering and Markov decision processes. Topics includes linear and logistic regression, regularization, probabilistic (Bayesian) inference, SVMs and kernel methods, ANNs, clustering, and dimensionality reduction. The course uses the Python programming language and assumes in addition familiarity with linear algebra, probability theory, and multivariate calculus. This course is designed to give graduate-level students a thorough grounding in the methodologies, technologies, mathematics and algorithms currently needed by people who apply machine learning to a whole host of applications.

Learning Objectives:

- Understanding a wide variety of learning algorithms.
- Develop skills to apply learning algorithms to solving practical problems.
- Understanding how to perform evaluation of learning algorithms and model selection.
- Implement in code common ML algorithms (as assessed by the homeworks).

Prerequisites:

Students in the class are expected to have a reasonable degree of mathematical sophistication, and to be familiar with the basic knowledge of linear/matrix algebra, multivariate calculus, probability and statistics. Undergraduate classes in these subjects should be sufficient. Students are also expected to have knowledge of basic algorithm design techniques (greedy, dynamic programming, randomized algorithms, linear programming, approximation algorithms) and basic data structures. Programming in Python is required.

Recommended Textbooks:

Bishop = "*Pattern Recognition and Machine Learning*", by C.M. Bishop, Springer, 2006

ESL = "*The Elements of Statistical Learning*", by T. Hastie, R. Tibshirani, and J. Friedman, Springer, 2008

Review Materials:

Linear Algebra: http://viterbi-web.usc.edu/~adamchik/567/review-linalg.pdf

Probability: http://viterbi-web.usc.edu/~adamchik/567/review-prob.pdf

Python Tutorial: http://cs231n.github.io/python-numpy-tutorial/

Google Colab: https://colab.research.google.com/notebooks/intro.ipynb

Theory Written Assignments:

- There will be four written theory assignments.
- The assignments should be submitted electronically via Desire2learn.
- Theory assignments must be neatly written or typed, for example in MS Word, and then converted to pdf.
- You may work in groups of 2-3. However, each person should hand-in their own writeup.
- Collaboration should be limited to talking about the problems, so that your writeup is written entirely by you and not copied from your partner.
- There are NO late days for assignments, and we will not accept late submissions.
- We won't regrade assignments.

Programming Assignments:

- There will be four programming (in Python) assignments.
- We advise you to use Google Colab interactive environment.
- Programming assignments should be submitted electronically to Desire2learn.
- Collaboration should be limited to talking about the problems.
- Each assignment will be checked for code plagiarism.
- There are NO late days for assignments, and we will not accept late submissions.

Exams:

- There will be two midterm online exams. The exam time is limited to the lecture time.
- The practice exams will be posted.
- Before each exam we will schedule a TA review session.
- No makeup exams will be provided.
- If you skip the second exam, you may be eligible for an IN grade for the course. The incomplete grade has to be completed within one year. However, in order to get an IN you have to have a valid cause. Please read the University policy on IN grade for more details.
- The exam solutions and grading rubric will always be posted.
- There will be a regrading session for each exam where you can discuss grading errors. A regrade is allowed only when there are clear and obvious grading errors. Grading errors are simple mistakes made on the part of the graders, and not differences in interpretation of a question or answer.

Piazza & Emails:

If you have a question about the material or logistics of the class, please do not use e-mail but instead post it on the Piazza at http://piazza.com/usc/summer2022/csci567

You may post it publicly to the whole class or privately to the instructors. Often times, if one student has a question/comment, other also have a similar question/comment. Please DO NOT send emails to the course staff unless your issue is private and/or a private post on Piazza is unsuitable.

`

Grading:

Theory assignments	**25%**
Programming assignments	**20%**
First midterm exam	**25%**
Second midterm exam	**30%**

Letter Grade Distribution:

≥ 90	A	60 – 65	C+
85 – 90	A-	55 – 60	C
80 – 85	B+	50 – 55	C-
73 – 80	B	45 – 50	D+
65 – 73	B-	≤ 45	D

Office Hours:

For the programming assignments:

Pranjal, Gautam	pranjal@usc.edu
Dash, Shubhashree	shubhash@usc.edu
Ghosal, Shuvam	sghosal@usc.edu

For the theory assignments and the lecture material:

TBA

Monday	Tuesday	Wednesday	Thursday	Friday

Schedule:

Each lecture is 2hrs 05 mins long. Lectures include a problem-solving (discussion) session.

Day	Date	Topics Covered
Wed	May 18	Lecture 1: Course Overview
Thu	May 19	Lecture 2: kNN, Cross-validation, Leave-one-out
Mon	May 23	Lecture 3: Decision Tree, Entropy and Gini impurity, Reduced-Error Pruning
Tue	May 24	Lecture 4: Linear Regression, Ridge and Lasso Regularizations
Wed	May 25	Lecture 5: Kernel Methods
Thu	May 26	Lecture 6: Support Vector Machines, Lagrangian
Mon	May 30	Memorial Day (no class)
Tue	May 31	Lecture 7: Lagrangian Duality, Dual SVM
Wed	June 1	Lecture 8: Perceptron, Logistic Regression, Surrogate Losses,
Thu	June 2	Lecture 9: Multiclass Classification, Gradient Descent
Mon	June 6	Lecture 10: Artificial Neural Network, Backpropagation
Tue	June 7	Lecture 11: Convolutional Neural Networks, R-CNN
Wed	June 8	Review for exam
Thu	June 9	Exam – I
Mon	June 13	Lecture 12: Ensemble Learning, Boosting, AdaBoost
Tue	June 14	Lecture 13: Generative Learning, Naïve Bayes
Wed	June 15	Lecture 14: Dimensionality Reduction, Principal Component Analysis
Thu	June 16	Lecture 15: K-means clustering, Kernel Density Estimation
Mon	June 20	Guest Lecture
Tue	June 21	Lecture 16: Gaussian Mixture Models, EM algorithm
Wed	June 22	Lecture 17: Hidden Markov Models
Thu	June 23	Lecture 18: Viterbi algorithm, Baum-Welch algorithm
Mon	June 27	Review for exam
Tue	June 28	Exam – II

Programming Assignments:

Assignment	Content	Out	Due
PA1	Decision Trees	May 19	May 27
PA2	SVM	May 27	June 7
PA3	CNN	June 7	June 15
PA4	RNN	June 15	June 22

Theory Assignments:

Assignment	Content	Out	Due
HW1	KNN, DT, Regression (lectures 2 - 4)	May 19	May 26
HW2	Kernels, SVM, Lagrangian (lectures 5 - 9)	May 26	June 6
HW3	Boosting, NB, PCA (lectures 12 - 14)	June 14	June 20
HW4	Clustering, GMM, HMM (lectures 15 - 18)	June 20	June 26

Disclaimer:

Although the instructor does not expect this syllabus to drastically change, he reserves every right to change this syllabus any time in the semester.

Academic Integrity:

The USC Student Conduct Code prohibits plagiarism. All USC students are responsible for reading and following the Student Conduct Code, which appears on https://policy.usc.edu/files/2018/07/SCampus-2018-19.pdf.

In this course we encourage students to study together. This includes discussing general strategies to be used on individual assignments. However, all work submitted for the class is to be done individually. Some examples of what is not allowed by the conduct code: copying all or part of someone else's work (by hand or by looking at others' files, either secretly or if shown), and submitting it as your own; giving another student in the class a copy of your assignment solution; consulting with another student during an exam. If you have questions about what is allowed, please discuss it with the instructor.

For Students with Disabilities:

Any student requesting academic accommodations based on a disability is required to register with Office of Student Accessibility Services (OSAS) each semester. A letter of verification for approved accommodations can be obtained from OSAS. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. OSAS is located in GFS 120 and is open 8:30 a.m.- 5:00 p.m., Monday through Friday.