Syllabus
New York University Tandon School of Engineering
Computer Science & Engineering
Course Outline CS-GY 6923 Machine Learning
Spring 2021
Raman Kannan

To contact professor:

Course Pre-requisites

Course Description

This course is an introduction to the field of machine learning, covering fundamental techniques for classification, regression, dimensionality reduction, clustering, and model selection. A broad range of algorithms will be covered, such as logistic regression, neural networks, deep learning, support vector machines, tree-based methods, expectation maximization, and principal components analysis. In addition meta-algorithms and other recent advances in M/L will be covered. The course will include hands-on exercises with real data from different application areas. Students will learn to train and validate machine learning models and analyze their performance.

Course Objectives

Students are expected to attain

1. conceptual understanding of both Supervised/Unsupervised Learning Techniques. Understand the statistical/algebraic foundation of these techniques, relative strengths and weakness, theoretical and practical criteria in adopting a model.
2. Understanding the process discipline: collect, describe, model, explore and verify data.
3. Engineering. Use industry standard environment and process to conduct repeatable and reproducible classification experiments.
4. Experimentation and Analysis: Run prescribed process to optimize model using multiple classification algorithms, evaluate them using standard performance metrics.
5. Deliver summary results of the experiments and explain key decisions they made in designing the model and model output.

Course Structure

This is an online course. All lectures, meetings are done using zoom accessed through newclasses.nyu.edu
We meet on Mondays at 8:30 PM.
Office hours on Mondays 4 to 6 PM by appointment via email by Sunday 5 PM.

For participation, students have to make something original and comment on two comments made by other students before the next class.

Readings

URL: https://statlearning.com/ISLR%20Seventh%20Printing.pdf

AUTHORS:Trevor Hastie, Robert Tibshirani, Jerome Friedman

TITLE:An Introduction to Statistical Learning

An optional and recommended text

URL:http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

AUTHORS:Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

TITLE:Elements of Statistical Learning, CODE:ESL

Online resources: https://www.quora.com/What-is-the-best-book-to-learn-ML

Stackoverflow and other sources

Green01: Required Reading

pages.stern.nyu.edu/~wgreene/Text/Greene-EA-7&8ed-Appendices.pdf

math.nyu.edu/~cfgranda/pages/DSGA1002_fall16/material/linear_algebra.pdf

https://www.cns.nyu.edu/~eero/math-tools/Handouts/linalg_jordan_86.pdf

Fisher01: Required Reading

Fisher's Discriminant Analysis: Linear Discriminant Analysis, Multivariate QDA [ISLR: 4.6.3,4.6.4]

NN:Neural Nets (any introductory material will suffice)

Chapter 11 intro to ML 3rd Edition, Ethem Alpaydin

Chapter 04 and Chapter 05 Miroslav-Kubat-Springer

Grade Distribution

Quality of Performance	**Letter Grade**	**Range %**
	A+	97-100
Excellent - work is of exceptional quality	A	93 – 96.9
	A-	90 - 92.9
Good - work is above average	B+	87 - 89.9
Satisfactory	B	83 - 86.9
Below Average	B-	80 - 82.9
Poor	C+	77 - 79.9

	C	70 - 76.9
Failure	F	< 70

Grade Calculation

Grades in this course are determined by the percentage of points obtained.

Course assignments	**Percentage of Final Grade**	**Points**
Homework – 1 (HW01)	30.00%	30
✓ First modeling assignment (15%) due 03/15/2021 ✓ Analyze, improve performance (15%) due 04/01/21 using Feature Selection or Penalization		
Class Discussion	12.00%	12
✓ You must participate in weekly forums and discussions. ✓ Discussions are applied analysis from the texts. ✓ You must post a response by Wednesday at midnight (ET) ✓ You must respond to at least one of your colleagues' contributions by Saturday at midnight (ET), ✓ You should provide meaningful feedback on the analysis.		
Homework – experiments with ensemble techniques	40.00%	40
✓ Improve your work in HW01 using CV/Bagging/Boosting/RandomForest/Stacking – 25% ✓ Due 05/10/2021 – analyze and summarize, compare and contrast these techniques and their utility to improve performance – 15%		
✓ Open book, open notes review test – t1 03/21/21	9.00%	9
✓ Open book, open notes review test – t2 04/25/21	9.00%	9
Total	**100%**	**100**

Course requirements

Submit all assignments before 11:59 PM on the due date, specified above.

Course Outline:

Please note that this schedule is subject to change depending on progress, questions, requests, etc.

Week	Topics	Tentative Date/Reading
1	Machine Learning, Supervised Learning (Classification) Unsupervised Learning(Clustering). Reinforcement Learning is not in scope. Supervised Learning:Generic Concepts applicable to all supervised learners: Occam's Razor,No Free Lunch Theorem, Induction/Generalization, Loss Function Minimization (aka Optimization), Bias/ Variance,Inability to learn, inability to generalize	02/01/2021 ISLR:Chap01, Chap 2.1,2.2.2, 2.2.3, Chap 4.1
2	Refresher Advanced Probability[Expected Value and Linear Algebra, Matrices, Vectors]/Statistics[i.i.d, CLT, LLN,descriptive/summary statistics,moments]/ Dataset manipulation in R, Datasets, Numerical/Categorical data, Scale and Models	02/08/2021 ISLR:Chap02 Greene01
3	Supervised Learning: Regression, Logistic Regression, Antidote to overfitting: regularization techniques: Ridge (L2), Lasso (L1) Classifier performance:TP,FP,TN,FN RoC, AUC, Accuracy,Specificity, Sensitivity,Precision, Recall,	No class on 02/15/2021 We meet on 02/17/2021 (Raman is n/a on 02/18/2021) 02/22/2021 ISLR:Chap 4.1, 4.3, 4.6.2
4	Sigmoid Function, activation function, perceptron (aka neural nets or ANNs) Extending the perceptrons with back propagation, hidden layers, other activation function Explainability, regularization	02/22/2021 ISLR:
5	Uncorrelated Features:Naïve Bayes Generative vs Discriminative Classifiers	03/01/2021
6	Instance Based techniques (no assumptions about the distribution, aka non-parametric) Distance, Nearest Neighbor kNN	03/08/2021 Chap 4.6.5
7	Curse of Dimensionality, Mahalanobis Distance, Dimensionality Reduction (LDA as Feature Selection, LASSO as Feature Selection, PCA (Cholesky,Eigen, SVD)	03/15/2021 ISLR:4.4.2 ISLR: 6.3
8	Decision Trees (no assumptions about the distribution, aka non-parametric), Entropy, Information Gain	03/22/2021 ISLR:8
9	Support Vectors (no assumptions about the distribution, aka non-parametric) Support Vector Machines: Margins, Kernel, Radial Basis, Gaussian	03/29/2021 ISLR:9
10	What causes inferior performance, techniques for performance improvement Resampling:Varying dataset Bootstrapping, Cross Validation, Bagging, Boosting	04/05/2021 ISLR:Chap 05 ISLR:8.2
11	Combining Classifiers: Aggregating variants of one classifier, combining heterogeneous Classifiers, Stacking	04/12/2021 ISLR:
12	Relevance of Stacking/CV/Bagging to parallelism and NFL	04/19/2021
13	Unsupervised Learning;Clustering Topic Modeling in Text Analytics,SOMs	04/26/2021 ISLR:Chap. 10

14	Semi Supervised – leveraging strengths of unsupervised and supervised	05/03/2021
15	Spring Final Exam	05/10/2021

05/10/2021 is the last day to submit HW02.

Moses Center Statement of Disability
If you are student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

NYU School of Engineering Policies and Procedures on Academic Misconduct – complete Student Code of Conduct here

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

    1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an

exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
3. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
4. Unauthorized collaboration: working together on work meant to be done individually.
5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

NYU School of Engineering Policies and Procedures on Excused Absences – complete policy here

A. Introduction: An absence can be excused if you have missed no more than **10 days of school**. If an illness or special circumstance has caused you to miss more than two weeks of school, please refer to the section labeled Medical Leave of Absence.

B. Students may request special accommodations for an absence to be excused in the following cases:

1. Medical reasons
2. Death in immediate family
3. Personal qualified emergencies (documentation must be provided)
4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the Coordinator of Student Advocacy, Compliance and Student Affairs and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

NYU School of Engineering Academic Calendar – complete list here.
The last day of the final exam period is _____. Final exam dates for undergraduate courses will not be determined until later in the semester. Final exams for graduate courses will be held on the last day of class during the week of _____. If you have two final exams at the same time, report the conflict to

your professors as soon as possible. Do not make any travel plans until the exam schedule is finalized.
Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc.
For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu