QBS 108: Applied Machine Learning

Spring Quarter 2020
Instructor: Saeed Hassanpour
TAs: Ben Ricard, Diana Song, David Chen with help from Behnaz Abdollahi
Class Time: Tuesday & Thursday, 4:30-6:00 pm

Course Description

This course provides a comprehensive introduction to machine learning methods and techniques. Various machine learning concepts and methods, such as natural language processing and deep learning, will be described and discussed. The emphasis of this course will be providing the required background and working knowledge of the machine learning methodology to apply these techniques on new or existing research or data science problems. Through multiple project assignments, this course will provide students with the experience on the application of machine learning techniques to solve real-world complex problems, such as those in the biomedical domain.

Course Learning Outcomes

In this course, students will:

- Understand the underlying bases of machine learning frameworks and methodologies
- Obtain a working knowledge of machine learning methods
- Become ready to develop machine learning solutions for real-world data science problems

Teaching Methods & Philosophy

Students are encouraged to read materials outside of class and ask questions in class. A teacher always learns from the students. All questions are a chance for both the questioner, other students, and the teacher to increase their understanding.

Class Climate & Inclusivity

At Dartmouth, we value integrity, responsibility, and respect for the rights and interests of others, all central to our Principles of Community. We are dedicated to establishing and maintaining a safe and inclusive campus where all have equal access to the educational and employment opportunities Dartmouth offers. We strive to promote an environment of sexual respect, safety, and well-being. In its policies and standards, Dartmouth demonstrates unequivocally that sexual assault, gender-based harassment, domestic violence, dating violence, and stalking are not tolerated in our community. These principles are also enumerated in the Professionalism Policy for the Undergraduate Medical Education Program at Geisel.

The Sexual Respect Website at Dartmouth provides a wealth of information on your right with regard to sexual respect and resources that are available to all in our community. Please note that, as a faculty member, I am obligated to share disclosures regarding conduct under Title IX with Dartmouth's Title IX Coordinator. Should you have any questions, please feel free to contact Dartmouth's Title IX Coordinator or the Deputy Title IX Coordinator for Geisel or for Guarini.

Prerequisites

- Intermediate Proficiency in Python
  - All homework assignments will be in Python.
- Calculus and Linear Algebra
  - Knowing how to take derivatives and understanding matrix/vector operations.
- Basic Knowledge of Probability and Statistics
  - Familiarity with probability distributions, mean, standard deviation, etc.

Suggested Reading

Alpaydin, Ethem. Introduction to machine learning. MIT press, 2014.

Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron C. Courville, Available online.

Assignments and Exams:

There will be 4 homework assignments for this class. Each homework includes multiple programming problems based on real-world datasets. Students are expected to submit the programming assignments in Python. The assignments will also include some non-programming questions on fundamentals of machine learning. The due date for each homework assignment is two weeks after the release. Students are allowed to have four late days for their assignment submissions in total.

The course has an in-class closed-book midterm and final. The questions in both exams are multiple choice and/or short answer questions.

Grade Distribution:

- Homework assignments: 60%
  - Each homework assignment: 15%
- Exams: 40%
  - Midterm: 20%
  - Final: 20%

Course Schedule and Topics

Session	Date	Topics	Demos	Homeworks
1.	03/31/20	Basic class information Introduction to artificial intelligence Logic-based systems Brute force systems	Demo on AI	Homework1 – Release
2.	04/02/20	Machine learning basics/Examples Supervised/Unsupervised /Semi Supervised Supervised learning pipeline Linear Regression	Demo of regression with visualization	
3.	04/07/20	Optimization using gradient descent Naïve Bayes theory Text classification Discrimination vs generative classifiers		
4.	04/09/20	Multinomial Logistic Regression Hyperparameter Tuning Regularization Classification Evaluation	Demo on LR (metrics training/validation/test sets) Demo on regularization and gradient descent	
5.	04/14/20	Decision trees Ensemble learning		Homework1 – Due Homework2 – Release
6.	04/16/20	Boosting Bagging Random forest kNN	Demo on KNN	
7.	04/21/20	Hierarchical clustering k-Means clustering Support vector machines	Demo on k-Means clustering	
8.	04/23/20	Dimensionality reduction Principal Component Analysis Singular Value Decomposition Feature selection/extraction	Demo on principle component analysis	
9.	04/28/20	Introduction to neural networks		Homework2 – Due
10.	04/30/20	Mid Term (In class - Covered material till 04/23/20)		
11.	05/05/20	Feed forward neural network Backpropagation Regularization, batch normalization, drop out		Homework3 – Release
12.	05/07/20	Optimization methods Training a neural network Transfer learning		
13.	05/12/20	Autoencoder tSNE Convolutional neural network Pooling Convnet		

14.	05/14/20	Important CNN architectures (AlexNet, VGG, GoogLeNet, Resnet, etc.) NLP introduction Language models Word representation BOW One-hot/tf/tf-idf encoding	
15.	05/19/20	Word2Vec Sequence Classifiers: RNN	Homework3 – Due Homework4 – Release
16.	05/21/20	LSTM Sequence to sequence classifiers: Machine translation Image captioning	
17.	05/26/20	Major NLP tasks Text classification Part of speech tagging Named-entity recognition Parsing IR	
18.	05/28/20	Variational Autoencoder GAN Reinforcement learning Research highlights, examples, and future directions of machine learning in biomedical applications	
19.	06/02/20	Guest Lecture: Building Trustworthy AI for Healthcare Amar Das, MD, PhD Program Director, IBM Research	Homework4 – Due
20.	06/04/20	Final exam (The same time and location as the class - Covered material after 04/23/20)	