COM-3920: Machine Learning

Learning Objectives

Course Level Learning Outcomes

Use and analyze existing learning algorithms, including methods for classification, regression, structured prediction, clustering, and representation learning

Integrate multiple facets of practical machine learning in a single system: data preprocessing, learning, regularization and model selection

Describe the formal properties of models and algorithms for learning and explain the practical implications of those results

Compare and contrast different paradigms for learning (supervised, unsupervised, etc.)

**Week 1**.

**Topic:** Overview of Machine Learning.

**Outcomes**

* Formulate a well-posed learning problem for a real-world task by identifying the task, performance measure, and training experience
* Describe common learning paradigms in terms of the type of data available and when, the form of prediction, and the structure of the output prediction
* Identify examples of the ethical responsibilities of an ML expert

**Questions**

* What Does it Mean to Learn?
  + *The general supervised approach to machine learning: a learning algorithm reads in training data and computes a learned function f.*
* What is the difference between memorization and generalization?

**Evidence**

* Ask students to formulate a learning problem (in class)
* Homework (focusing on revisions and assessing what they know in terms of basic probability and statistics, calculus, and linear algebra)

**Week 2**.

**Topic: Linear Regression**

* + To know simple linear regression, intercept slope coefficient parameter, least squares, sum of squares, population regression line, bias and unbiased, standard error, residual standard error
  + Design k-NN Regression and Decision Tree Regression
  + To Understand how to estimate the regression coefficients
  + Implement learning for Linear Regression using three optimization techniques: (1) closed form, (2) gradient descent, (3) stochastic gradient descent
  + To be able to choose a Linear Regression optimization technique that is appropriate for a particular dataset by analyzing the tradeoff of computational complexity vs. convergence speed

**Questions**

* How does KNN work and what are the pros and cons of KNN? How can we tune the hyperparameter K?
* What are the differences between the KNN classifier and KNN regression methods?
* How do we choose the step-size and convergence criteria?
* How do we measure the quality of fit?
* Is There a Relationship Between the Response and Predictors?

**Evidence**

* Quiz on KNN and Regression
* Case study: Predicting Housing Prices
* Homework
  + Problems
  + Applying concepts on a dataset (Python)