PHYS 243: Foundation of Applied Machine Learning: An Empirical Approach

Bahram Mobasher

TA : Abtin Shahidi

Summer, 2019

TA E-mail: abtin.shahidi@email.ucr.edu
Website: **https://abtinshahidi.github.io/teaching/2019-summer-foundation-machine-learning**
Github page: **https://github.com/abtinshahidi/Foundation_applied_machine_learning**
Slack group: Invitation link

Course Description

Machine learning has emerged as a very powerful tool in Data Science. With ML techniques, computational systems can adaptively improve their performances using experimental data to train. This allows construction of algorithms that can learn from and make predictions based on data. Machine learning is a demanding discipline and is presently used to extract information from data in a variety of fields. This course is designed to prepare students to work in the Data Science disciplines using ML techniques, introduce existing ML algorithms and develop such algorithms. The course will provide practical experience and case studies based on real data. It covers examples from different disciplines- physics, astronomy, biology, neuroscience, finance etc. The course also covers an introduction on deep learning. The course will provide an extensive introduction to the current graduate course in ML (CS229) and will complement that course. A practical and hands-on course in machine and deep learning is not currently offered at UCR. During this course guest lecturers will be invited to give lectures on practical aspects of machine learning

Required Materials

- Course notes and jupyter notebooks are available on the course website and github page

Prerequisites/Corequisites

Prerequisites: This course does not assume any specific technical background, but we will use linear algebra and calculus as well as basic python programming.

Course Objectives

Successful students:

1. Ability to gain insight by looking at a given dataset
2. Ability to clean and preprocess any given dataset
3. Ability to write probabilistic programs
4. Ability to run Monte Carlo simulations
5. Ability to model the hypothesis using different algorithms
6. Ability to implement minimalist version of different algorithm. (Particularly learning algorithms)
7. Ability to choose a proper model and to evaluate the model
8. Ability to use important libraries such as `numpy`, `scipy`, `matplotlib`, `pandas`, `skit-learn`, `Keras`, `tensorflow`

Grading Policy

- **25%** of your grade will be determined by **Homework assignments** (5% each).
- **25%** of your grade will be determined by **Midterm project**
- **25%** of your grade will be determined by **Written Final exam**
- **25%** of your grade will be determined by **Final project**

Course Policies

Policies on Incomplete Grades and Late Assignments

For every day passed after the deadline you will lose 10 percent of the assignment grade. Basically you won't get any points after 10 days

Academic Integrity and Honesty

Students are required to comply with the university policy on academic integrity found in the academic integrity at UCR stated policies.

Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide at the end of the semester.

Week 01, 06/24 - 06/28: General definitions, and python programming

- Background and Introduction
- Historical Development of machine learning Applications of Machine Learning What we learn in this course Python Programming- installation on Window or Mac platforms. (Python 3)
- Getting familiar with Bayesian statistics. (link)

Week 02, 07/01 - 07/05: Bayesian statistics

- Getting familiar with Bayesian statistics. (link)

Week 03, 07/08 - 07/12: Basic linear algebra, Statistics, Simulations

- Basic linear algebra
- Review of statistics and probability theory
- Matrix formulation and calculation of least squares Multivariate Calculus
- Partial Derivative Gradients
- Monte Carlo Simulation (link)

Week 04, 07/15 - 07/19: Multivariate statistics , Statistics, Confidence intervals

- Multivariate statistics
- Joint distributions Mean vectors
- variance-covariance matrices
- Conditional distributions
- marginal distributions Multivariate normal distributions and their basic properties
- Confidence intervals (link)

Week 05, 07/22 - 07/26: Machine Learning: Regression, Clustering and Classification

- Classifications
- K-Nearest Neighbors
- Logistic regression
- Clustering and classification implementation (link)

Week 06, 07/29 - 08/02: Machine Learning: Regression, Gradient Descent, Decision Trees

- Regressions Linear and polynomial Regressions
- Ridge Regression
- Over-fitting
- Regularization
- Decision Trees, Gradient Descent, Stochastic Gradient Descent (SGD) Regression (link)

Week 07, 08/05 - 08/09: Decision Trees, Random Forests, Support Vector Machines

- Support Vector Machine (link)
- Cross validation
- Receiver Operating Characteristic (ROC) curves P-R Curves
- Principle Component Analysis (PCA)
- Decision Trees, Random Forests, Support Vector Machines implementation

Week 08, 08/12 - 08/16: Neural networks, extras

- Neural Nets and Artificial Neural Nets
- Deep Learning
- Convolutional Neural Nets
- autoencoders
- Multi-layer networks Back propagation

Week 09, 08/19 - 08/23: Manifold Learning, Generative models/ Review

- Dimensional reductions SOM, t-SNE, LLE
- Adversarial Neural Nets, Recurrent Neural Nets