Course title: Physics 6820/8820, Big Data Analytics in Physics
Fall 2019
Tuesday/Thursday 12:40pm - 2:45pm
Instructor: Richard Hughes

This course provides an introduction to machine learning and advanced algorithms, with an emphasis (as much as possible) on practical physics-based applications, using publicly available data sets. The goal is to provide an introduction to Data Science for students who may want to pursue this as a career option and/or apply these techniques in a research environment.

Students are expected to come into the course with the following:
0. A laptop/chromebook or similar. Mac/Linux/Windows/ChromeOS are all fine, since we will be using a browser-based environment (Jupiter notebooks) for all of our programming.
1. Basic programming skills, which could be in any of a number of different languages, such as C++, java, python, etc. All course assignments will be done in python. Entering into the course after doing a simple online python tutorial would be a good idea!
2. Basic knowledge of statistics and probability (such as would be obtained in Physics 4700).
3. Enthusiasm for learning and a desire to challenge oneself!

Grading: Grading is based on the following:
a) 70%: In class and out of class assignments
b) 30%: Final Project

Syllabus:

Week	Topic
1	Python Intro/Python challenge; Manipulating data files; Visualization with matplotlib and plotly
2	Intro to linear regression; Dealing with missing data; test/train splits; feature scaling and categorical data
3	Intro to Classification using support vector machines; The confusion matrix; ROC curves and AUC
4	Multi-class classification;

	k-fold validation
5	Decision Trees and Random Forests; Over- and Under-fitting, and the Bias-Variance Tradeoff; Feature Importance
6	Linear Regression and Gradient Descent; writing your own regressor from scratch Logistic Regression (write your own from scratch)
7	Using the Ohio SuperComputer Center (OSC) batch system; Softmax Regression Neural Networks from scratch
8	Introduction to Keras: The Industry Standard Neural Network Library Multi-Layer Perceptrons
9	Convolutional Neural Networks
10	Project Proposal Due Autoencoders; Stacked Autoencoders and Classification
11	Visualization of learned Features in Neural Networks Adversarial Examples Variational Autoencoders
12	1D Convolutional Neural Networks and Sequences Text Classification Project Progess Report Due
13	Siamese Networks and the iPhone Face Recognition Algorithm
14	Possible Additional topics: Generative Adversarial Networks, Recurrent Neural Networks
15	During Exam Week: Project due (no final exam)