Course Description

Machine learning uses interdisciplinary techniques such as statistics, linear algebra, optimization, and computer science to create automated systems capable of sifting through large volumes of data to predict or make decisions without human intervention. Machine learning is now extremely widespread, with applications ranging from business intelligence to homeland security, from the study of biochemical processes to structural analysis, pollution to astrophysics, and more. This class will familiarize students with a wide cross-section of machine learning models and algorithms, and prepare students to apply machine learning techniques to research or industry.

Prerequisite:

Either all of the following CPE 101, CPE 202, Math 244, Math 241 (or equivalent) OR CPF Python. Everyone is welcome, however, not having the prerequisites could make this class much harder.

Overview:

Based on fundamental knowledge of computer science principles and skills, probability and statistics theory, and the theory and application of linear algebra. This course provides a broad introduction to machine learning.

Instructor, Time, and Location:

Instructors: Hari Krishna, Shaheen Alemi. Time and Location: TBD

Office Hours:

By appointment. Other office hours are to be determined. Contact: hkrish01@calpoly.edu

Textbook: None required. Recommended: Deep Learning, Goodfellow et al.

Learning Outcomes:

By the end of the course, students should be able to:

- Develop an appreciation for what is involved in learning models from data.
- Understand a wide variety of learning algorithms.
- Understand how to evaluate models generated from data.
- Apply the algorithms to a real-world problem, optimize the models learned and report on
- the expected accuracy that can be achieved by applying the models.

Note: This is **not** an easy class/workshop. It is fast paced and covers multiple topics that would otherwise take multiple classes. It is less rigorous, but more material.

Course Schedule

Day 1, Week 1	Data Science Intro
Day 2, Week 1	Cleaning, Manipulating, Collecting, and Visualizing Data
Day 3, Week 2	Linear Regression
Day 4, Week 2	Logistic Regression
Day 5, Week 3	Decision trees, forests, bias/variance tradeoff
Day 6, Week 3	Bayes and Other ML (learning theory)
Day 7, Week 4	Support Vector Machines
Day 8, Week 4	Neural Networks
Day 9, Week 5	Intro to Big Data and SQL
Day 10, Week 5	Dimensionality Reduction and Feature Selection
Day 11, Week 6	Clustering and Bagging and Boosting
Day 12, Week 6	Convolutional Neural Networks
Day 13, Week 7	Computer Vision
Day 14, Week 7	NLP and LSTM
Day 15, Week 8	Reinforcement Learning
Day 16, Week 8	Special Topics: Autonomous Driving and Bioinformatics

Optional Homework:

We can't force you to do homework and some of you won't like homework, however, computer science is a lab based discipline. Work with teams, alone, or not do it at all, you get out of this course how much you want to put in.

All materials to study will be on the Github including textbooks, homework, labs, and slides.

Prerequisite Exam (not necessary, just to gauge readiness):

Question 1:

Calculate the gradient of the following function:

f(x,y) = $x^2 - 3y^2x$

Question 2:

What is the relationship between angles of vectors and their dot products?

Question 3:

What are the dimensions of the following matrix multiplication:

$$\begin{pmatrix}1&1&1&1&1&1&1\\1&1&1&1&1&1&1\end{pmatrix}\cdot\begin{pmatrix}1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\\1&1&1&1&1&1&1\end{pmatrix}\cdot\begin{pmatrix}1&1\\1&1\\1&1\\1&1\\1&1\\1&1\\1&1\end{pmatrix}$$

dimensions: 2x7 * 7x7 * 7x2

Question 4:

What does the following code do:

```
import numpy as np
a = np.arrange(50).reshape((5,10))
```

Question 5:

What is the interpretation of the following coefficient of b:

y = 6a + 7b + 82