**Machine learning**

Instructor: Dr. Vikas Ramachandra

Topics to be covered:

1 Linear regression

1. Simple Linear Regression
2. Hypothesis Testing
3. Multiple Linear Regression
4. Model Selection
5. Interactions and Non-Linear Models

2. Classification

1. Introduction
2. Logistic Regression
3. Multivariate Logistic Regression
4. Multiclass Logistic Regression
5. Linear Discriminant Analysis
6. Univariate Linear Discriminant Analysis
7. Multivariate Linear Discriminant Analysis
8. Quadratic Discriminant Analysis

3. Resampling methods: Cross Validation and bootstrap

1. Prediction Error and Validation Set
2. K-Fold Cross-Validation
3. Cross-Validation Do's and Don'ts
4. Bootstrap

4. Variable Selection

1. Linear Model Subset Selection
2. Forward Stepwise Selection
3. Backward Stepwise Selection
4. Estimating Test Error -- Mallow's Cp, AIC, BIC, Adjusted R-squared
5. Estimating Test Error -- Cross-Validation
6. Ridge Regression
7. Lasso
8. Tuning Parameters
9. Dimension Reduction
10. Principal Components and Partial Least Squares

5. Non-Linear Models

1. Polynomial Regression
2. Piecewise Regression and Splines
3. Smoothing Splines
4. Local Regression and Generalized Additive Models

6. Decision Tree models

1. Decision Trees
2. Pruning Trees
3. Classification Trees
4. Bootstrap Aggregation (Bagging) and Random Forests
5. Boosting

7. Support Vector Machines

1. Maximal Margin Classifier
2. Support Vector Classifier
3. Kernels and Support Vector Machines
4. Comparison with Logistic Regression

8. Unsupervised learning: Principal Components and Clustering

1. Principal Components Analysis
2. Proportion of Variance Explained
3. K-Means Clustering
4. Hierarchical Clustering
5. Example of Hierarchical Clustering

Lec. 9-15: TBD

Possible topics:

Bayesian techniques, generative models, neural networks and deep learning, reinforcement learning, graphical models, data manipulation (munging), exploratory data analysis, end-to-end ML model pipeline+deployment using R/Shiny.

Pre-requisites and prior reading:

Basic probability and linear algebra topics, basic programming concepts.

Book and resources:

Introduction to statistical learning

<http://www-bcf.usc.edu/~gareth/ISL/>

This is a free ebook with additional resources on the above webpage.

It provides an excellent introduction to machine learning.

We will be following this book for about half of the course, and other resources as we proceed to more advanced topics in the few final lectures of the course. Additional resources will be provided to you weekly as the class progresses.

Homework, projects and assignments:

Homework:

Please complete watching the lab videos from this link:

<http://fs2.american.edu/alberto/www/analytics/ISLRLectures.html>

Weekly, for each corresponding lesson.

Also, please write code to mimic the code in the videos, preferably on your own datasets,

And ideally, a different dataset for each exercise.

There are several interesting datasets available online, please choose what you find interesting.

You are expected to turn in the code, as well as results (such as an R markdown document at the very least).

Graded projects:

There will be 4 graded projects, you are free to pick datasets, from the real world.

You are expected to try multiple machine learning methods taught in this course (or beyond, based on your interest), for each project. You are also expected to do pre-processing of the dataset, some exploratory plots, fit ML models, and evaluate them against your chosen metrics.

We expect to see both your code, as well as a presentation explaining your process, results, etc. Time permitting, there will also be an oral examination to test your understanding of concepts, which will also serve as mock interviews to mimic data science job interviews.