NATIONAL UNIVERSITY OF SINGAPORE
NUS Business School
Department of Decision Sciences

DSC5103 Statistics

Lecturer : Assistant Prof Wang Tong

Session : Semester I, 2015/2016

Aims & Objectives

This course aims to provide a holistic overview of the modern Statistical Learning toolbox. Different from traditional Statistics courses, this course (1) emphasizes on understanding the intuition behind the tools and not on deriving the underlying mathematics; (2) incorporates real-world datasets and analytics projects to help you bridge theories and practices; and (3) equips you with hands-on experiences in using data analysis software (R and Python) to visualize the concepts and ideas and also solve exercises.

The course covers most of the commonly used analytics tools such as logistic regression and decision tree. Two exceptions, Support Vector Machine and Neural Network (which are arguably more of Computer Science tools), are left to other modules.

The students are expected to get their hands really dirty by applying (and even messing up) the tools in analytics software (R or Python).

Topics

Week 1. Overview of Statistical Learning
1. Descriptive, Predictive, and Prescriptive Analytics
2. Supervised, Unsupervised, and Reinforcement Learning
3. Classification and Regression
4. Bias-Variance Trade-off, Under-fitting vs. Over-fitting
5. Simple K-nearest Neighbors Algorithm

Week 2--4. Review of Classic Statistics Concepts
1. Random Variables, Basic Probability Distributions
2. Confidence Interval, Hypothesis Testing, P-value
3. Maximum Likelihood Estimation and the Expectation-Maximization Algorithm
4. Time Series Analysis (Exponential Smoothing, ARIMA)

Week 5. Linear Regression
1. Simple and Multiple Linear Regression
2. Interpreting Regression Output
3. Introducing interactions and nonlinearity

Week 6. Model Selection and Resampling Methods
1. Linear Model Selection
2. Cross-validation
3. The Bootstrap

Week 7. Generalizations of Linear Regression
1. Logistic Regression
2. Poisson Regression and other Generalized Linear Models

Week 8. Regularization
1. Ridge Regression
2. The Lasso

Week 9. Tree-based Methods I: Decision Trees

Week 10. Tree-based Methods II:
1. Bagging, Random Forest
2. Boosting Machines

Week 11. Unsupervised Learning
1. K-Means Clustering
2. Dimension Reduction by Principle Component Analysis

Week 12. Review and Miscellaneous Topics

Text Book

An Introduction to Statistical Learning – with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, 2013, Springer-Verlag New York.

http://www-bcf.usc.edu/~gareth/ISL/index.html

Reference Book (for those who want more math)

The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed, by Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2009, Springer-Verlag New York.

http://statweb.stanford.edu/~tibs/ElemStatLearn/

ASSESSMENT (100% CA)

Class Participation (Individual)	10%
Assignments (Group)	30%
In-class Tests (Individual)	40%
Final Project (Group)	20%