INFO 6105

Data Science Engineering Methods

Exam Syllabus

The exam will be in two parts.

1. A full class in class exam near the end of the semester
2. An hour online HackerRank exam in our finals slot.

TOPICS:

Everything in the book. There will be some questions from the conceptual questions in the back of the book.

An Introduction to Statistical Learning with Applications in R (2013)

Authors: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

Free online via SpringerLink (http://link.Springer.com/) <http://link.Springer.com/book/10.1007/978-1-4614-7138-7>

Specifically,

Understanding Probability Distributions

Random variables, quantiles, mean variance

Central limit theorem

Bias versus variance

Significance tests and standard scores

Resampling methods: bootstrapping

Unsupervised Learning (Clustering)

Selecting and Evaluating Models

Validation

Cross-Validation

ROC curves

Regression and classification metrics

Regression

Logistic Regression

Regularization

Linear Models, Model Inference and Interpretation

Dummy variables, multicollinearity, interaction effects

Supervised Learning

Decision trees

Random Forests

Support Vector Machines

Bagging, Boosting and Stacking

K-Means Clustering

Hierarchical Clustering