

CSCI 447 Exam 1 Outline

- I. Machine Learning Process
 - A. Gathering evidence
 - B. Modifying the model
 - C. Evaluating effectiveness
- II. Machine Learning Dimensions
 - A. Input
 - 1. Features
 - 2. Data
 - B. Algorithms
 - C. Output
 - D. Evaluation
 - E. Other
- III. Probability
 - A. Random Variables
 - B. Joint and Marginal Distributions
 - C. Conditional Distributions
 - D. Product Rule, Chain Rule, Bayes Rule
 - E. Independence
- IV. Linear Algebra
 - A. What is Linear?
 - B. Inputs and Operations
 - C. Operations as Inputs
- V. Python / AWS
 - A. Data (S3 Buckets)
 - B. Processing (in AWS)
 - C. Output (Evaluating Results)
- VI. Data Preparation
 - A. Dataset Issues
 - 1. Missing Data
 - 2. Erroneous Data
 - 3. Data Format
 - B. Multiple Source Issues
 - 1. Different Features
 - 2. Different Answers
 - 3. Population Sample
- VII. Linear Regression
 - A. Linear Models
 - B. 1D Ordinary Least Squares (OLS)
 - C. Solution of OLS
 - D. Interpretation
 - E. Anscombe's Quartet
 - F. Multivariate OLS
 - G. OLS Pros and Cons
- VIII. Logistic Regression
 - A. Math Behind Logistic Regression

- B. Visualizing Logistic Regression
- C. Loss Function
- D. Batch / Full Regression
- E. Gradient Descent
 - 1. OLS
 - 2. Logistic Regression
- F. Comparing OLS and Logistic Regression
- G. Multi-Class Logistic Regression
- IX. Summing Up Regression
 - A. Assumptions / Extensions of Linear
 - B. Beyond Linearity
 - C. Extreme Learning Machines
 - D. Overfitting
 - E. Regularization
 - F. Cross Validation
- X. Clustering
 - A. Hierarchical
 - B. K-Means
- XI. Nearest Neighbor
 - A. K-Nearest Neighbor