

CS 381/780 Data Analytics Final Exam Review Topics

- SQL
 - join tables, various aggregate functions
- Know your Statistics
 - Hypothesis Testing, Type I vs Type II errors
 - Skews and Kurtosis and Correlation
 - Correlation and Causation
- General Machine Learning principle
 - Bias vs Variance, overfit vs underfit
- Linear Regression
 - R-squared, Adjusted-R squared, MSE, Cost function, Gradient descent
- Logistics Regression
 - Confusion matrix, Precision vs Recall, True Positive Rate, False Positive Rate
 - Sigmoid function, Odd vs Probability, Principle of maximum Likelihood

CS 381 Data Analytics Final Exam Review Topics

- Support Vector Machine
 - Margin, Soft Margin, Hyperplane
 - Support Vectors, Principle of Maximum Margin
 - C-parameters (bias and variance trade-off)
 - Kernel trick
 - Pros and Cons when compared with other models
- Decision Trees
 - Gini index, Entropy, Information gain
 - How to choose which attribute to split
 - Max_depth
- Random Forest
 - Ensemble techniques

CS 381 Data Analytics Final Exam Review Topics

- Clustering
 - K-means algorithms
 - Elbow graph
- Naïve Bayes Classifier
 - Bayes Theorem
- Text Mining
 - Stopwords
 - Term Frequency, Inverse Document Frequency
 - Term by Document Matrix
- Big Data
 - What is the problem, and what is the solution
 - Advantages of using the Cloud