

# INTRO TO DATA SCIENCE RIEVIEW



---

## WHAT YOU HAVE DONE 1

---

Exploratory Data Analysis

python

numpy

pandas

matplotlib

Gradient Descent

Normal Equations

Generalized Linear Models

Linear Models for Linear Regression

Linear Models for Non-linear Problems

Logistic Regression, including multi-class

classification

Principal Components Analysis

K-means Clustering

Naive Bayes Models

Bernoulli

Multinomial

Gaussian

Headline Testing

Natural Language Processing

Count Vectorization

TF-IDF

Cosine Similarity

Latent Dirichlet Allocation - Topic Modeling

Recommendation Systems

Collaborative Filtering

K-nearest Neighbors

Support Vector Machines

Non-linear and linear

Regression and classification

Ensemble Methods

Averaging

Boosting

Decision Trees

MapReduce and Spark

Time Series Analysis

Neural Networks

---

## WHAT YOU HAVE DONE 2

---

Feature selection

Feature scaling

Polynomial Features

Validation and cross validation

Regularization

Bias

Variance

Mean Squared Error

Accuracy

Precision

Recall

F1-Score

ROC

AUC

Gaussian Distribution

Beta Distribution

Bernoulli Trial

Multinomial Distribution

Dirichlet Distribution

Grid Search

Random Grid Search

---

## DATASETS

---

Chicago House Prices

IMDB

Africa Soil

Grateful Dead

Baseball Hitters

Challenger Disaster

Iris Dataset

20 News groups

UN Countries

Citibike Data

MNIST

Movie Critics

Tweet Sentiments

Airlines Dataset

Book-Crossing Dataset

Engine Misfiring Dataset

Libor Dataset

Ebola Dataset

---

## WHAT HAVEN'T WE TOUCHED ON?

---

Bayesian Regression

Bayesian Classification (Logistic Regression)

Discriminant Functions - Fisher's Linear Discriminant

Only touched on Neural Networks

Didn't investigate Convolutional Neural Networks,

Restricted Boltzman Machines, Deep Belief Networks,

Recurrent Networks, Generative Models

Radial Basis Function Networks

Bayesian Networks

Gaussian Processes

Expectation Maximization

Mixture of Gaussians

Graphical Models & Approximate Inference

Sampling Methods

Probabilistic PCA

Factor Analysis

Independent Component Analysis

Auto-associative Neural Networks and Non-linear

PCA

Bayesian PCA

Advanced Optimization Methods - Conjugate

Gradients

Markov Modeling

Hidden Markov Modeling

---

**...AND FINALLY**

---

On behalf of myself, Susan and Chris thank you

We wish you the best of luck in an future Data Science Endeavors

Please feel free to send me a linkedin invite