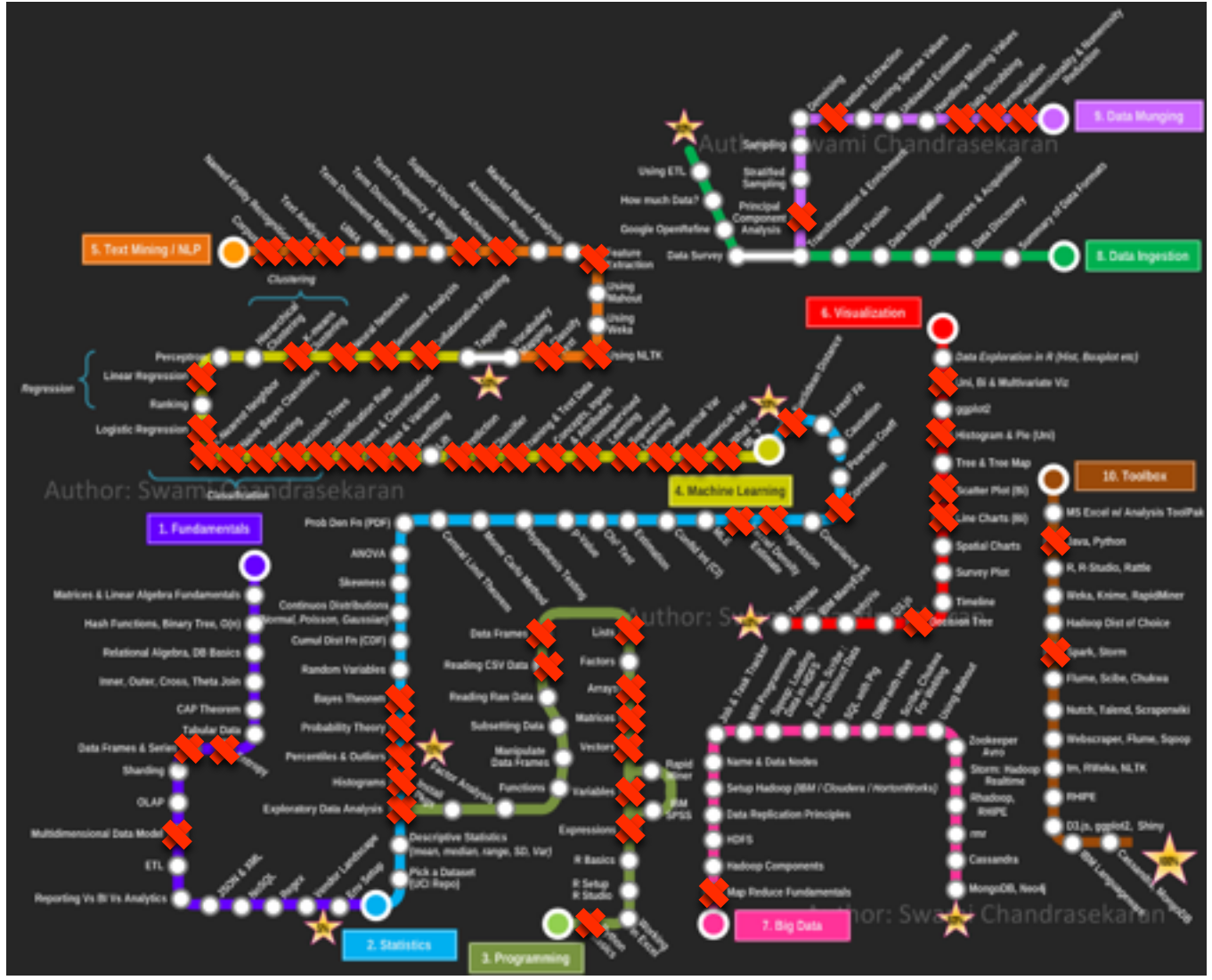
GENERAL ASSEMBLY

INTRO TO DATA SCIENCE RIEVIEW


1. Fundamentals
2. Statistics
3. Programming
4. Machine Learning
5. Text Mining / NLP
6. Visualization
7. Big Data
8. Data Ingestion
9. Data Munging
10. Toolbox
Author: Swami Chandrasekaran

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

Gaussian Processes
Expectation Maximization
Mixture of Gaussians
Graphical Models & Approximate Inference
Sampling Methods
Probabilisitic PCA
Factor Analysis
Independent Component Analysis
Bayesian PCA

Advanced Optimization Methods - Conjugate Gradients
Markov Modeling
Hidden Markov Modeling

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
Auto-associative Neural Networks and Non-linear PCA

…and Finally

On behalf of myself, Susan and Chris thank you

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

Please feel free to send me a Linkedin invite