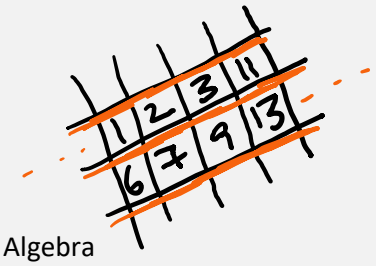
Data Science Learning Path

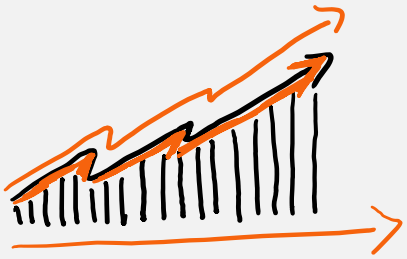
Right from basics to complex problem-solving level. Sign up for multiple courses at the same time and do them just as you would in a university – My suggestion is to take 2-3 items for a 4 months period. Following are the topics comprising of the whole Data Science journey – Under each main topic are several important concepts you must cover, along with good reference material from the most widely chosen education platforms. Happy Learning!

Algebra and Matrices



- Linear Equations and Graphs
- Functions
- Transformations
- **Reference Material**
  - **Coursera** - Mathematics for Machine Learning: Linear Algebra
  - **edX** – Linear Algebra I - IV by Georgia Tech
  - **Book** – Elementary Linear Algebra by Stephen Andrilli and David Hecker

Statistics – I

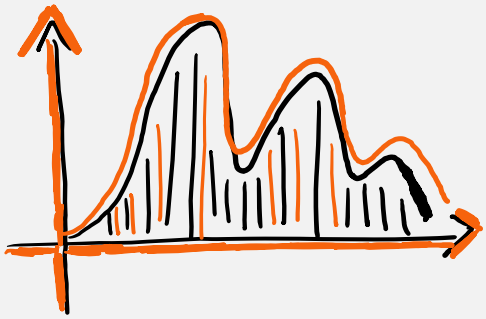


- Probability
- Modelling and Sampling data distributions
- Mean, Median, Mode, Variance, Standard Deviation
- Confidence Intervals
- **Reference Material**
  - **Udacity** – Statistics by San Jose State University
  - **edX** – Probability and Statistics I- IV by Georgia Tech
  - **Book** - Introduction to Probability and Statistics Principles and Applications for Engineering and the Computing Sciences by J. Milton and J. Arnold

Analytical Thinking

- Inference Building
- Reading and understanding data
- Correlations and Similarities
- **Reference Material**
  - **Udacity** – Intro to Inferential Statistics
  - **Coursera** – Statistical Inference by John Hopkins University
  - **edX** – Statistical Thinking for Data Science and Analytics by Columbia University

Calculus



- Differentiation
- Integration
- Chain Rule
- Optimization
- **Reference Material**
  - **Coursera** - Mathematics for Machine Learning: Multivariate Calculus by Imperial College of London
  - **Coursera** – Calculus and Optimization for Machine Learning by NRUHSE
  - **Book** – Thomas' Calculus

Programming – Concepts and Implementation

- Introduction to programming
- Variables, Functions, Loops
- Object Oriented Principles and Design
- Introduction to Data Structures: Arrays and Lists
- **Reference Material**
  - **Book** – Structure and Interpretation of Computer Programs by Harold Abelson, Gerald Jay Sussman, Julie Sussman
  - **edX** – CS50's Introduction to Computer Science by Harvard University
  - **Coursera** – Object Oriented Programming in Java Specialization by Duke University
  - **edX** – Introduction to Python Programming by Georgia Tech
  - **Coursera** – Python 3 Programming by University of Michigan
  - **Coursera** – R Programming by John Hopkins University
  - **Udacity** - Object Oriented Programming in Java
  - **Book** – The Pragmatic Programmer: Your Journey to Mastery by Andrew Hunt, David Thomas

Data Structures and Algorithms

- Introduction to Algorithms
- Concepts: Sorting, Searching, Divide and Conquer, Shortest Path, Greedy Algorithms
- Data Structures Implementation: Arrays, Lists, Queues, Stack
- Algorithmic Complexities and Constraints
- Advanced Data Structures and Algorithms: Trees, Graphs, String-based, Special case-based
- **Reference Material**
  - **Udacity** – Data Structures and Algorithms
  - **Coursera** – Data Structures and Algorithms Specialization by UC San Diego
  - **Coursera** – Algorithms Part 1 and 2 by Princeton University
  - **edX** – Algorithms and Data Structures by UC San Diego
  - **Book** - Introduction to Algorithms is a book on computer programming by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein
  - **Book** – Algorithms by Robert Sedgewick

Database Management

- Introduction to Database and DBMS
- Data Models: Terminology and Rules
- SQL: Query Formulation
- Data Models: Entity Relationship Diagrams
- Data Models: Schema Building
- Normalization
- **Reference Material**
  - **Coursera** – Introduction to Structured Query Language by University of Michigan
  - **edX** – Databases I – V by Stanford University
  - **Book** – Database Systems – The Complete Book by Hector G. Molina, Jeffrey D. Ullman, and Jennifer Widom

Python / R for Machine Learning

- Python / R Programming Fundamentals
- Advanced Python: Data Structures and Comprehensions
- Advanced R: R Functions, Debugging and Profiling
- Focus Python Libraries: Numpy, Pandas, Matplotlib
- Focus R packages: Dplyr, DT, Caret
- **Reference Material**
  - **edX** – Python Basics for Data Science by IBM
  - **edX** – Python for Data Science by UC San Diego
  - **Coursera** – IBM Data Science Professional Certificate courses 1 – 5
  - **Coursera** – Introduction to Data Science in Python by University of Michigan
  - **Coursera** – Data Science: Foundations using R Specialization by John Hopkins University

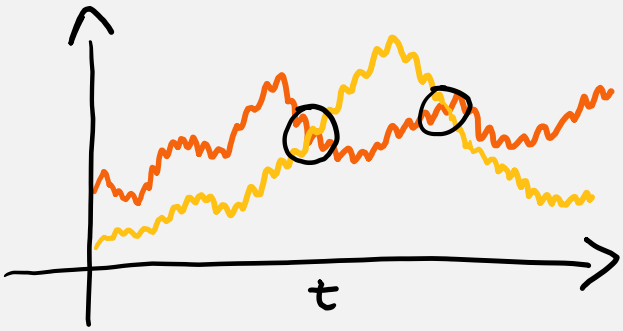
Exploratory Data Analysis

- Introduction to Data Analysis
- Statistical Inference
- Data Exploration: Forming Insights
- **Reference Material**
  - **Udacity** – Intro to Descriptive Statistics
  - **Coursera** – Developing Data Products by John Hopkins University
  - **Coursera** – Data Analysis with Python by IBM
  - **Book** – Exploratory Data Analysis by John Tukey

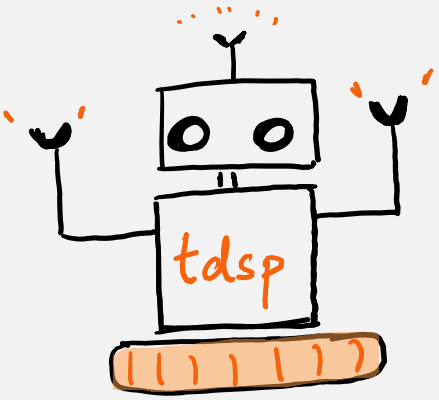
Data Visualization

- Visualization Best Practises
- Focus Python Libraries: Matplotlib, Seaborn, Bokeh
- Focus R packages: Ggplot2, Esquisse, Shiny
- **Reference Material**
  - **Coursera** – Data Visualization with Python by IBM
  - **edX** – Data Science: Visualization by Harvard University
  - **Book** – Storytelling with Data: Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic

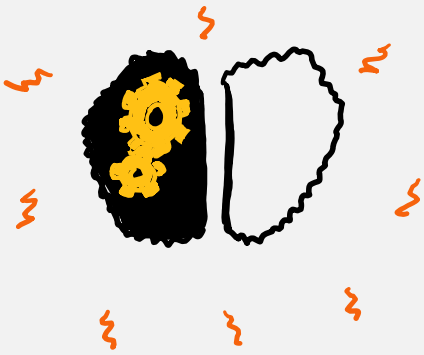
Statistics – II



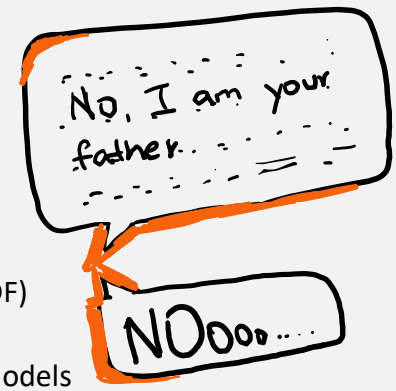
- Linear Data Modelling
- Multivariate Analysis
- Regression Analysis
- Time Series Analysis
- Factor and Path Analysis
- **Reference Material**
  - **Book** – Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
  - **Book** – The Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani, Jerome Friedman



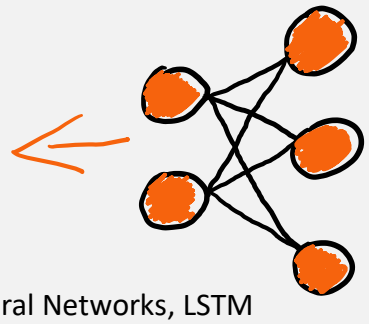
- **Artificial Intelligence**
  - Problem Solving with Search Algorithms
  - Probability with Bayes Rule and Bayesian Network
  - Knowledge Representation, Logic
  - Planning
  - Constraint Satisfaction
  - Introduction to Machine Learning
  - Introduction to Natural Language Processing
  - **Reference Material**
    - **edX** – Artificial Intelligence by Columbia University
    - **edX** – Artificial Intelligence by Georgia Tech
    - **edX** – CS50's Introduction to Artificial Intelligence with Python
    - **Udacity** – Intro to Artificial Intelligence
    - **Book** – Artificial Intelligence: A Modern Approach by Peter Norvig, Stuart J. Russell



- **Machine Leaning – I**
  - Basics of Machine Learning
  - Types of Machine Learning Algorithms
  - Supervised v/s Unsupervised Learning
  - Cost Function and Regularization
  - Optimization Algorithms
  - Classification v/s Regression
  - Perceptron and Logistic Regression
  - Different Machine Learning Algorithms
  - Dimensionality Reduction
  - Working with different types of data and underlying challenges
  - Introduction to Deep Learning
  - **Reference Material**
    - **edX** – Machine Learning by Georgia Tech
    - **edX** – Machine Learning by Columbia University
    - **Coursera** – Machine Learning with Python by IBM
    - **Coursera** – Machine Learning by Stanford University



- **Natural Language Processing**
  - Basics of Natural Language Processing
  - Stop word elimination, Stemming, Lemmatization, Regex
  - Feature Extraction and Encoding from Text
  - Term Frequency (TF) and Inverse Document Frequency (TF-IDF)
  - Word Embeddings
  - Vector Space Models, N-gram Language Models, Sequence Models
  - Attention Models
  - Applications: Sentiment Analysis and Machine Translation
  - **Reference Material**
    - **Coursera** – Natural Language Processing Specialization by deeplearning.ai
    - **edX** – NLP and NLU by University of Texas Arlington



- **Deep Learning**
  - Basics of Neural Network and Perceptron Model
  - Neural Network Design and Architecture
  - Parameter and Hyperparameter Tuning
  - Regularization and Optimization
  - Advanced: Convolutional Neural Networks, Recurrent Neural Networks, LSTM
  - **Reference Material**
    - **Coursera** – Deep Learning Specialization by deeplearning.ai
    - **Coursera** – Applied Text Mining in Python by University of Michigan
    - **edX** – Deep Learning by IBM
    - **Book** – Deep Learning by Aaron Courville, Ian Goodfellow, Yoshua Bengio

- **Machine Learning – II**
  - Machine Learning Approach and Process
  - TensorFlow Development
  - Advanced Algorithms: Recommender Systems, Reinforcement Learning
  - Structuring Machine Learning Projects
  - Deploying Machine Learning Models
  - **Reference Material**
    - **Coursera** – Advanced Machine Learning Specialization by NRUHSE
    - **Coursera** – Reinforcement Learning Specialization by University of Alberta
    - **Coursera** – Recommender Systems Specialization by University of Minnesota
    - **Coursera** – TensorFlow Developer by deeplearning.ai
    - **Coursera** – TensorFlow: Data and Deployment by deeplearning.ai

- **Machine Learning Problems**
  - **Kaggle** – Datasets and Competitions – Practise!

- **Wrapping up**
  - **Reference Material**
    - **Coursera** – Data Science Specialization by John Hopkins University
    - **Coursera** – Data Warehousing for Business Intelligence Specialization by University of Colorado

- **Specific software and platforms**
  - **Reference Material**
    - **edX** – Introduction to Data Analysis using Excel by Microsoft
    - **edX** – Analyzing and Visualizing Data with Excel by Microsoft
    - **edX** – Analyzing and Visualizing Data with Power BI by Microsoft
    - **Coursera** – Data Visualization with Tableau by UC Davis
    - **Coursera** – Data Engineering, Big Data and Machine Learning on GCP Specialization
    - **dataiku** - Academy