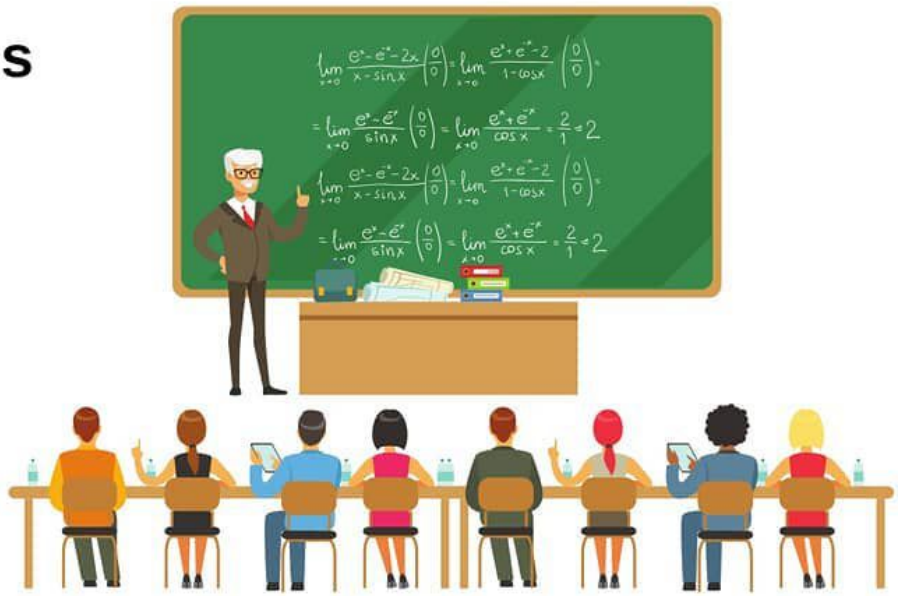
GUIDE TO BECOMING A SELF-TAUGHT DATA SCIENTIST 2020

Math Basics

1. **Multivariable Calculus**
2. **Functions of several variables**
3. **Derivatives and gradients**
4. **Step function, Sigmoid function, Logit function, ReLU (Rectified Linear Unit) function**
5. **Cost function**
6. **Plotting of functions**
7. **Minimum and Maximum values of a function**

1. **Linear Algebra**
2. **VectorsMatrices**
3. **Transpose of a matrix**
4. **The inverse of a matrix**
5. **The determinant of a matrix**
6. **Dotproduct**
7. **Eigenvalues**
8. **Eigenvectors**

Math Basics



@learn.machinelearning

1. **Probability and Statistics Basics**
2. Mean, Median, Mode,
3. Standard deviation / variance
4. Correlation coefficient and the covariance
5. matrixProbability distributions (Binomial, Poisson, Normal)
6. p-valueBaye's Theorem Confusion Matrix, ROC Curve)
7. A/B Testing
8. Monte Carlo Simulation

1. **Optimization Methods**
2. Cost function/Objective function
3. Likelihood function
4. Error function
5. Gradient Descent Algorithm and its variants (e.g., Stochastic Gradient Descent Algorithm)

Programming Basics

1. **R**
2. **Basic R syntax**
3. **Foundational R programming concepts such as data types, vectors arithmetic, indexing, and data frames**
4. **How to perform operations in R including sorting, data wrangling using dplyr, and data visualization with ggplot2**
5. **R studio**

1. **PYTHON**
2. **Basic Python syntax**
3. **Object-oriented programming**
4. **Jupyter notebooks**
5. **Python libraries such as**
6. **NumPy, pylab, seaborn**
7. **matplotlib, pandas**
8. **scikit-learn**
9. **TensorFlow**
10. **PyTorch**
11. **etc.....**

Learn Data Basics

1. Learn how to manipulate data in various formats, for example, CSV file, pdf file, text file, etc.
2. Learn how to clean data, impute data, scale data, import and export data, and scrap data from the internet.
3. Some packages of interest are pandas, NumPy, pdf tools, stringr, etc.
4. Additionally, R and Python contain several inbuilt datasets that can be used for practice.
5. Learn data transformation and dimensionality reduction techniques such as covariance matrix plot, principal component analysis (PCA), and linear discriminant analysis (LDA).

Data Visualization Basics

1. Data Component
2. Geometric Component
3. Mapping Component
4. Scale Component
5. Labels Component
6. Ethical Component

Machine learning basics

1. **Supervised Learning (Continuous Variable Prediction)**
2. **Basic regression**
3. **Multi regression analysis**
4. **Regularized regression**
5. **Logistic Regression Classifier**
6. **Support Vector Machine (SVM)**
7. **K-nearest neighbor (KNN) Classifier**
8. **Decision Tree Classifier**
9. **Random Forest Classifier**
10. **Naive Bayesc)**
11. **Gradient boosting**
12. **etc .....**

1. **Unsupervised Learning**
2. **Kmeans clustering algorithm**
3. **k - Median**
4. **DBScan**
5. **Hierarchical clustering**
6. **etc...**

Practice

Form a team and practice what you learned on these platforms

Analytics Vidhya
Learn Everything About Analytics

kaggle

@learn.machinelearning

TECHGIG

DRIVENDATA

hackerearth

CrowdANALYTIX

CodaLab

Build up your Online Presence

- Write blogs
- Do projects an upload them on GitHub
  - Fork interesting repos
  - commit to other repos
- public speaking
- Youtube tutorials
- Share your experience on social channels.
- write books
- creat a course
- Twitch stream, or podcast.

Networking

- Make friends
- Meet experts and talk with them
- Learn from experts
- Get a mentor
- Make yourself visible to outside world
- It also helps you to get a good job in your dream companies

@learn.machinelearning

And then you are a

@learn.machinelearning

Thank You.

Like, Comment, Share and Save it for Later

Happy Learning

@learn.machinelearning