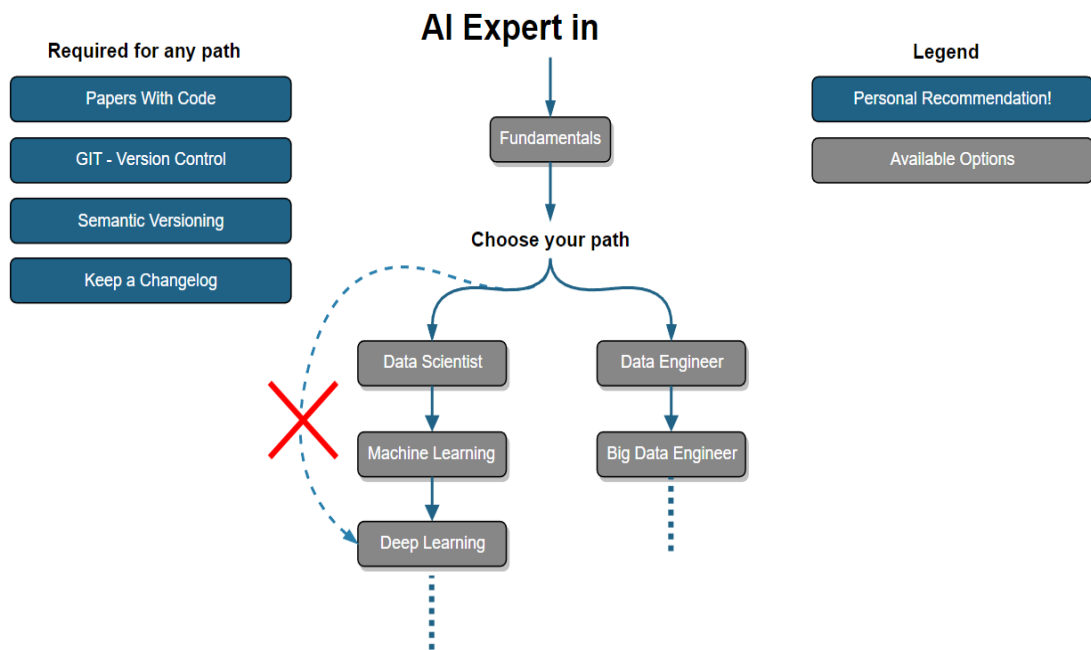
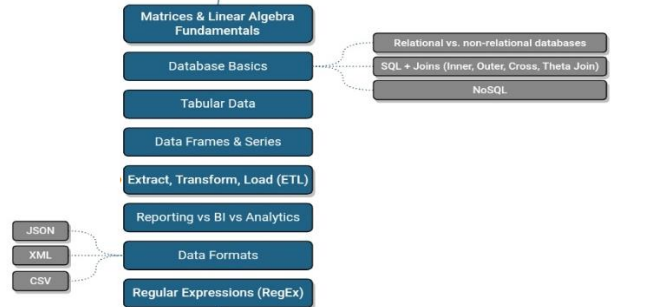


AI ROADMAP:

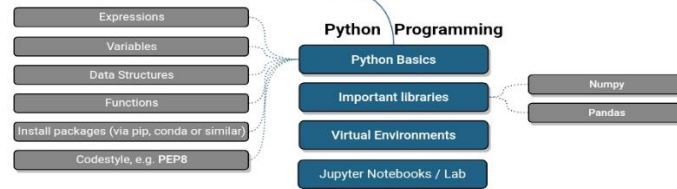


Fundamentals

Basics



Python Programming



Data Sources

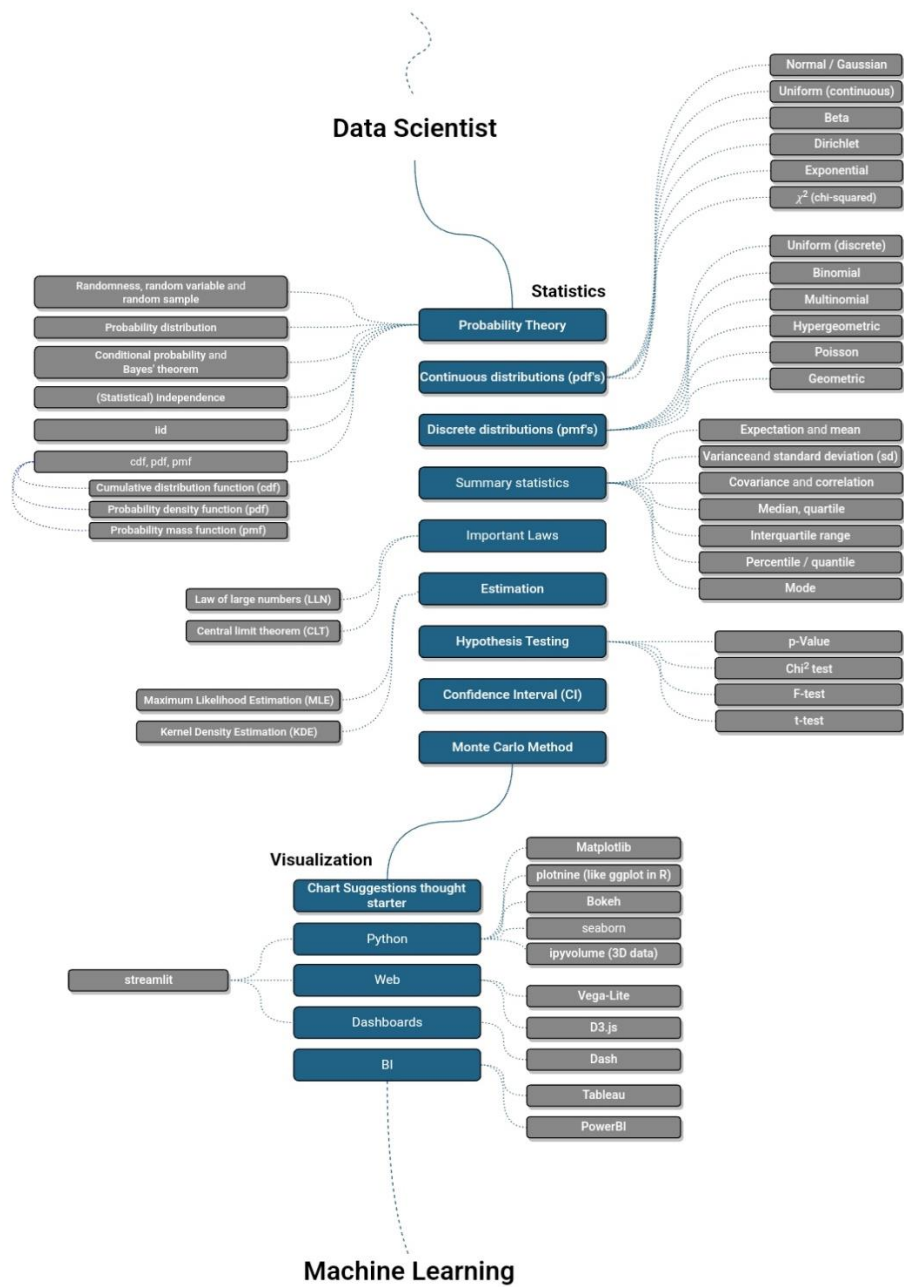


Exploratory Data Analysis / Data Munging / - Wrangling



Data Scientist

Data Engineer



Machine Learning

General

Concepts, Inputs & Attributes

Categorical Variables

Ordinal Variables

Numerical Variables

Cost functions and
gradient descent

Overfitting / Underfitting

Training, validation
and test data

Precision vs Recall

Bias & Variance

Lift

Methods

Supervised Learning

Unsupervised Learning

Ensemble Learning

Reinforcement Learning

Regression

Linear Regression

Poisson Regression

Classification

Classification Rate

Decision Trees

Logistic Regression

Naive Bayes Classifiers

K-Nearest Neighbour

SVM

Gaussian Mixture Models

Use Cases

Sentiment Analysis

Collaborative Filtering

Tagging

Prediction

Tools

Important libraries

scikit-learn

spacy (NLP)

Deep Learning

Clustering

Hierarchical Clustering

K-Means Clustering

DBSCAN

HDBSCAN

Fuzzy C-Means

Mean Shift

Agglomerative

OPTICS

Association Rule Learning

Apriori Algorithm

ECLAT algorithm

FP Trees

Dimensionality Reduction

Principal Component Analysis (PCA)

Random Projection

NMF

T-SNE

UMAP

Boosting

Bagging

Stacking

Deep Learning

Papers

- Deep Learning Papers Reading Roadmap
- Papers with code
- Papers with code - state of the art

Neural Networks

- Understanding Neural Networks
- Loss Functions
- Activation Functions
- Weight Initialization
- Vanishing / Exploding Gradient Problem

Architectures

- Feedforward neural network
- Autoencoder
- Convolutional Neural Network (CNN)
 - Pooling
- Recurrent Neural Network (RNN)
 - LSTM
 - GRU
- Transformer
 - Encoder
 - Decoder
 - Attention
- Siamese Network
- Generative Adversarial Network (GAN)
- Evolving Architectures / NEAT
- Residual Connections

Training

- Optimizers
 - SGD
 - Momentum
 - Adam
 - AdaGrad
 - AdaDelta
 - Nadam
 - RMSProp
- Learning Rate Schedule
- Batch Normalization
- Batch Size Effects
- Regularization
 - Early Stopping
 - Dropout
 - Parameter Penalties
 - Data Augmentation
 - Adversarial Training
- Multitask Learning
- Transfer Learning
- Curriculum Learning

Tools

- Important Libraries
 - Awesome Deep Learning
 - Huggingface Transformers
- Tensorflow
- PyTorch
- Tensorboard
- MLFlow

Model optimization (advanced)

- Distillation
- Quantization
- Neural Architecture Search (NAS)

Keep exploring and stay up-to-date



Data Engineer

Summary of Data Formats

Data Discovery

Data Source & Acquisition

Data Integration

Data Fusion

Transformation & Enrichment

Data Survey

OpenRefine

How much Data

Using ETL

Data Lake vs Data Warehouse

Dockerize your Python
Application

keep exploring and
stay up-to-date

Big Data Engineer

Big Data Architectures

Architectural Patterns & Best Practices (video)

Principles

- Horizontal vs vertical scaling
- Map Reduce
- Data Replication
- Name & Data Nodes
- Job & Task Tracker

Tools

Check the Awesome Big Data List

Hadoop (large data)

Spark (in memory)

RAPIDS (on GPU)

Flume, Scribe: For Unstruct Data

Data Warehouse with Hive

Elastic (EKL) Stack

Avro

Flink

Dask

Numba

Onnx

OpenVino

MLFlow

Kafka & KSQL

Databases

Scalability

Cloud Services

Awesome Production ML

HDFS

Loading data with Sqoop and Pig

Storm: Hadoop Realtime

to get data (e.g. logging), search, analyze and visualize it in realtime

Cassandra

MongoDB, Neo4j

AWS SageMaker

Google ML Engine

Microsoft Azure Machine Learning Studio

ZooKeeper

Kubernetes

keep exploring and stay up-to-date