Data Science: From Theory to Practice

Week 1 (Unit 1 – Basics):

Session 1: Introduction, review essential parts of Python (function, data collections, OO…)

Session 2: Anaconda setup, Jupyter notebook, Pandas, numpy

Project: Bartender project

Week 2:

Session 1: Statistics, Probability Distributions, Hypothesis Testing, t-test, p-value

Session 2: Visualizations (matplotlib, seaborn), ANOVA, MANOVA, F-test

Week 3:

Session 1: : A/B test, A/A test, Parametric & non-Parametric

Session 2: Bias and variance, analysis strategy

Project: Analysis report (using statistical models)

Week 4 (Unit 2 – Supervised Learning)

Session 1: Data cleaning, Data Exploration, Feature Engineering, Multicollinearity

Session 2: Principal Component Analysis (PCA), Data Transformation, Feature Selection

Assignment: PCA computation (by hand)

Week 5:

Session 1: Linear Regression, Multivariate Regression, Residue, Homoscedasticity, Maximum Likelihood Estimation

Session 2: Holdout, Cross Validation, Overfitting, Class Imbalance, Error Types (I &II), Partial Least Square Regression (PLSR), Gradient Descent Algorithm

Project: Melbourne Housing

Week 6:

Session 1: KNN, Naïve Bayes

Project: Amazon Reviews

Session 2: Decision Tree, Entropy, ID3 Algorithm, Pruning, Random Forest, Gradient Boosting

Assignment: Constructing Decision Tree (by hand)

Week 7:

Session 1: Logistic Regression, Ridge Regression, Lasso Regression, feature selection using RFE, SelectKBest

Assignment: calculating beta coefficients for Ridge & Lasso (by hand)

Session 2: Support Vector Machine (SVM), Kernel Trick, Ada Boosting, Stochastic Gradient Boosting

Assignment: Constructing separating line in SVM (by hand)

Project: Airline Arrivals

Week 8 (Unit 3 – Unsupervised Learning):

Session 1: K-means, Clustering, Cluster Evaluation, Silhouette score

Session 2: Customer segmentation persona and characteristics

Project: Boston Marathon

Week 9 (Unit 4 – Deep Learning):

Session 1: Neural Network, Back Propagation

Assignment: Construct 2-layers ANN and calculate adaptive weight (by hand)

Session 2: Deep Learning, CNN, Hidden Layer, Convolutional layer, Maxpooling layer, Sub-sampling layer

Week 10 :

Session 1: TensorFlow, Keras, Pytorch

Assignment: Constructing CNN (by hand)

Session 2: CNN architecture

Project: Fruit classification project

Week 11:

Session 1: RNN, LSTM

Session 2: work on fruit classification project

Week 12 (Unit 5 – Natural Language Processing):

Session 1: Bag of Words (BoW), BoW Features, supervised technique

Session 2: TF-IDF, Lament Semantic Analysis (LSA), Sentence Similarity

Assignment: calculate TF-IDF from text (by hand)

Week 13:

Session 1: Word2Vec, sense2vec, n-grams

Session 2: pLSA, Latent Dirichlet Allocation (LDA)

Project: Thousand texts – author classification

Week 14 (Unit 6 – Computer Vision):

Session 1: OpenCV, Processing Video, Tracking

Assignment: Face Recognition using Deep Learning

Session 2: work on face recognition project

Week 15 & 16: Final Capstone project

ADDITIONAL TOPICS

Week 17 (Unit 7 – Data Scraping):

Session 1: Scrapy

Session 2: API, Json, HTML scraping

Project: Temperature

Week 18 (Unit 8 – machine learning with big data):

Session 1: Big data, Hadoop

Session 2: Distributed Computing and Sparks

Week 19:

Session 1: Time Series, Stochastic Modeling, ARIMA

Session 2: Auto Regression, ARMA

Project: Stock price

Week 20:

Session 1: Markov Processes, Hidden Markov Random Field

Session 2: Restricted Boltzmann Machine, Autoencoders

Project: Feature Image Extraction