Data Science Syllabus - ITC

October 2018 Cohort

Summary (about 1 month each)

1. Fundamentals
2. Data Science Fundamentals and Modeling
3. Neural Networks, Computer Vision, and NLP
4. Workshops and Special Topics

Details

Fundamentals

The gist: core skills that are fundamental in most DS workflows.

This section of the program includes all the core components that are essential to practicing Data Science in industry, from good coding practices and data visualization techniques, to working in docker container, virtual environments and with databases, and more.

Emphasis will also be placed later in the course on how to 'put models in production' and engineering aspects around it. You will also learn the core components of Statistics and Linear Algebra that are pertinent to understand the "behind the scenes" of many models that you'll learn later on in the course.

Data Science Fundamentals and Modeling

The gist: modeling techniques and tools most commonly used in DS (not Neural Networks based)

In this section, you will learn the most important core techniques in Machine Learning and Data Science. Most of those techniques and algorithms do not involve Neural Networks, but are often simpler and better choices than NNs for many problems commonly found in industry.

In the Data Modelling module, you will be introduced to some of the most important concepts in Data Science and Machine Learning: under/overfitting, risk and loss functions, bias variance trade-off and model selection, regularization, cross validation, different metrics, and more. In this

module you will also learn about Naive Bayes, Logistic Regression, SVMs, Decision Trees, and Ensembles.

In addition, you will have classes on Hidden Markov Models, Time Series Analysis, some clustering techniques, and many matrix methods (like SVD, PCA, ICA, etc) and graph/network methods. These are all incredibly powerful tools when understood well and can help you i) understand your data, ii) process your data prior to feeding it another algorithm, or iii) straight-out solve the problem at hand.

In the Spark class, you'll learn about one of the most up-and-coming technologies in Data Science today. It allows, among other things, to handle data in a tabular format (like Pandas) at any scale and with great speed. Among the topics covered in the course will be how it enables you to interact easily with HDFS and S3. You'll also be exposed to different ways to deal with large datasets.

Deep Learning, Computer Vision, and NLP

The gist: the nuts and bolts of Neural Networks, Computer Vision, and Natural Language Processing

You will start with the fundamentals of Deep Learning and Neural Networks, which includes the core components that are relevant for the most common Deep Learning applications (such as in computer vision, natural language processing, time series, and more). Some of the topics covered here are backpropagation, common optimizers used and their differences, Tensorflow and Keras, different loss functions, types of layers, regularization techniques used for NNs, and more.

Tracks: Note that you will learn the fundamentals of both tracks, regardless of which you're in.

Computer Vision Track: In the Computer Vision Classics module, you will be exposed to the classic Computer Vision techniques (i.e. not DL based), such as feature extraction, convolutions, and pyramides and practice with openCV. These are still used in many applications, are useful on their own right, and also crucial for properly appreciating the following section on Deep Learning for Computer Vision. You will then learn Deep Learning applications specifically focused on vision tasks. These will include Convolution Neural Networks and modern architectures. This will bring you close to the state-of-the-art in these fields, and enable you to understand the main paradigms behind the famous recent papers and buzz.

Natural Language Processing Track: In the Natural Language Processing you will start from learning how to build a classifier with using one-hot vectorization and linear models and Support

Vector Machines. You'll then move toward word embeddings, and language models and into layers relevant for NLP, such as LSTM and GRU, and how to use RNNs for entity extraction. You will get practice and exposure to Spacy, NLTK, as well as the relevant tools in Keras and Tensorflow, and through exercises and other workshops get practice with a variety of settings for which NLP is useful.

Workshops and Special Topics

The gist: workshops with companies to learn more techniques and get a better sense of real world challenges.

Examples of past workshops:

- Recommender Systems: Outbrain, Taboola
- Autonomous Vehicles: Mobileye
- Cyber: Deep Instinct and Check Point
- Feature selection and engineering: SparkBeyond
- Information Retrieval: Yahoo, Sefaria
- GPUs and DL Solutions: NVIDIA
- Word Embeddings: SimilarWeb
- Medical Imaging: Aidoc
- Text Classification and Topic Modelling: Chorus