Advanced Data Science Bootcamp

Data Science Training Program
Booz | Allen | Hamilton
X Data Society

Building Scalable Models in Spark

8 hours of instruction

Learn how to optimize your code and to speed up current data processing using Spark. In this course, you will work through best practices of how and when to use Spark. You will explore what they can do with Spark and how to use distributed computing within Spark. Finally, you'll deploy a Spark application on AWS.

Syllabus & Topics Covered

1. Introduction to Spark and PySpark

- Working with data in PySpark
  - *RDDs vs. DataFrames vs. Datasets*
  - *When to use RDDs over DataFrames & Datasets*
- Features of Dataset in PySpark
  - *Optimized queries with Datasets*
  - *API for Datasets*

2. Introduction and overview of AWS for data science

- Overview the Data Analytics ecosystem within AWS
  - *Analytics*
  - *Data movement*
  - *Data lake*
  - *Predictive analytics and machine learning*

3. Working with ec2 in AWS

- Introduction and application of working with ec2
- Working with GPU using P2 instances

Prerequisites

Students must have experience with Python or another object-oriented language and foundational understanding of Spark.

AI: Neural Networks and Deep Learning

24 hours of instruction (Python)

Become an expert data scientist by implementing advanced techniques such as deep learning and neural networks. These state-of-the-art methods will allow you to build powerful predictive systems and find latent patterns in large amounts of data. This course dives into a host of algorithms known as artificial neural networks. You will learn the foundations of this complex and exciting topic, acquire practical skills to implement neural networks and deep learning models using TensorFlow and Keras libraries in order to solve real-world problems.

Syllabus & Topics Covered

1. Introducing the concept of neural networks

- Environment setup: TensorFlow 2.0
- Why TensorFlow?
- Introduction to neural networks
- Implement a neural network in Python using sklearn

2. Feed forward networks

- Forward and back propagation
- Single layer perceptron
- Implementation of models in TF

3. Best practices of model building – deep learning

- Batch learning
- GPU and when it is needed
- Data normalization
- Accuracy, precision, recall, F1 against "loss"

4. Auto encoders and Image classification - CNN

- Autoencoders and feature reduction
- Overview of CNNs and image data
- CNN architecture
- Simple CNNs – build and implement
- Optimizing CNN

Syllabus & Topics Covered Cont.

5. Recurrent Neural Networks (RNNs)

- RNN theory
- RNNs for time series data
- RNN implementation in TensorFlow
- Discuss best practices for time-series RNN models

6. Long short-term memory (LSTMs)

- Long Short-Term Memory (LSTM) theory
- LSTMs for text data
- LSTM implementation in TensorFlow

7. Reinforcement learning

- Reinforcement learning (RL) theory
- Implement RL using TensorFlow
- Expand CNNs to accomplish RNN tasks

Prerequisites

Any student in this program must have a strong foundation in Python and the common libraries: SciKit-Learn, Pandas, Numpy, and Matplotlib. Additionally, students should have a foundational knowledge of statistics, unsupervised machine learning algorithms, and classification algorithms.

Schedule

Tuesdays and Thursdays from 5pm - 7pm

Pricing

Payment: $1,450

Submit Your FlexEd Application

Apply for FlexEd Funding

You can apply for FlexEd funding. More info can be found here.

FlexEd Step-by-Step Guide

Use these FlexEd Instructions for a step-by-step guide to your application. The instructions and more key resources can be located in Degreed.

FlexEd Questions?

Visit the FlexEd FAQs here.