

Deep Learning with Databricks

Code:	DEEPLARNING
Length:	2 days
URL:	View Online

This course begins by covering the basics of neural networks and the tensorflow.keras API. We will focus on how to leverage Spark to scale our models, including distributed training, hyperparameter tuning, and inference, while leveraging MLflow to track, version, and manage these models. We will deep dive into distributed deep learning, including hands-on examples to compare and contrast various techniques for distributed data preparation, including Petastorm and TFRecord, as well as distributed training techniques such as Horovod and spark-tensorflow-distributor. To better understand the model's predictions, you will apply model interpretability libraries. Further, you will learn the concepts behind Convolutional Neural Networks (CNNs) and transfer learning, and apply them to solve image classification tasks. We will wrap up the course by covering Recurrent Neural Networks (RNNs) and attention-based models for natural language processing (NLP) applications.

Skills Gained

- Build deep learning models using tensorflow.keras
- Tune hyperparameters at scale with Hyperopt and Spark
- Track, version, and manage experiments using MLflow
- Perform distributed inference at scale using pandas UDFs
- Scale and train distributed deep learning models using Horovod
- Apply model interpretability libraries, such as SHAP, to understand model predictions
- Use CNNs and transfer learning for image classification tasks
- Use RNNs, attention-based models, and transfer learning for NLP tasks

Prerequisites

- Intermediate experience with Python and pandas (or completion of Introduction to Python for Data Science & Data Engineering)
- Familiarity with Apache Spark (or completion of Apache Spark Programming)
- Working knowledge of machine learning and data science (or completion of Scalable Machine Learning with Apache Spark)

Course Details

Course Outline

Course Outline

Day 1

- Neural network and tf.keras fundamentals
- Improve models by adding data standardization, callbacks, checkpointing, etc.
- Track and version models with MLflow
- Distributed inference with pandas UDFs
- Distributed hyperparameter tuning with Hyperopt
- Large scale data preparation with Petastorm

Day 2

- Distributed model training with Horovod and Petastorm
- Model interpretability with SHAP
- CNNs for image classification and transfer learning
- Distributed training with TFRecord using spark-tensorflow-distributor
- Deploy REST endpoint using MLflow Model Serving on Databricks
- Textual embeddings, RNNs, attention-based models, and transfer learning for named entity recognition (NER)

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy

Get Your Free Copy Now