

# Machine Learning Technology

*Tuesday 11 July 2017 13:30 (1h 30m)*

Machine learning (ML) is a thriving field with active research topics. It has found numerous practical applications in natural language processing, understanding of speech and images as well as fundamental sciences. ML approaches are capable of replicating and often surpassing the accuracy of hypothesis driven first-principles simulations and can provide new insights to a research problem.

This session will introduce machine learning technology focusing on the open source software stack built around TensorFlow and Apache Spark frameworks.

- Brief introduction to TensorFlow architecture and the primitives, implementing fully connected and convolutional layers, deep dive into higher-level APIs including `tf.layers`, estimators and Keras.
- Learn to debug machine learning applications and visualize training and cross validation process with TensorBoard. Hands-on demo: debugging convolutional neural net. Discuss ways to train multi-GPU and distributed models on a cluster
- Introduction to Spark transformations, actions, loading data into RDDs, DataFrames and Datasets, writing user-defined functions (UDF, UDAF). Discuss how to use Spark ML: transformers, estimators, pipeline. Creating your own UnaryTransformer

All exercises will use a mix of TensorFlow (Python API), and PySpark, Spark ML (parts of Apache Spark). Python programming experience is desirable, but previous experience with Tensorflow, Spark or distributed computing is not required.

**Presenter:** SVYATKOVSKIY, Alexey (Princeton University)