

Scalable Machine Learning 1:30 - 5:00pm

Machine learning is an integral part of knowledge discovery in a wide variety of applications. From scientific domains to social media analytics, the data that needs to be analyzed has become massive and complex. This session provides an introduction to approaches that can be used to perform machine learning at scale. Tools and procedures for executing machine learning techniques on HPC will be presented. Spark will also be covered. In particular, we will use Spark's machine learning library, MLlib, to demonstrate how distributed computing can be used to provide scalable machine learning. Please note: Knowledge of fundamental machine learning algorithms and techniques is required. (See description for Machine Learning Overview.)

1:30 - 2:15: Intro to R; R in HPC

2:15 - 3:15: Spark, MLlib

3:15 - 3:30: Break

3:30 - 4:45: Deep Learning

4:45 - 5:00: Wrapup