



Overview

Week 1

Week 2

Week 3

Week 4

Grades

Notes

Discussion Forums

Messages

Course Info

Week 3

Scalable Machine Learning on Big Data using Apache Spark

Week 3

Discuss this week's modules here.
227 threads · Last post 4 days ago

Go to forum

Week 3: Introduction to Apache SparkML



Romeo Kienzler

Understand the concept of machine learning pipelines in order to understand how Apache SparkML works programmatically

Key Concepts

- Summarize properties of machine learning pipelines
- Express machine learning pipelines programmatically using Apache SparkML
- Examine the importance of data preparation
- Explain clustering and the k-means algorithm
- Analyze any BigData data set using k-means and Apache SparkML

⬆ Less

Introduction to Apache SparkML



Video: How ML Pipelines work 3 min

Resume



Video: Introduction to SparkML 20 min



Video: Extract - Transform - Load 3 min



Reading: Exercise 1: Modifying a Apache SparkML Feature Engineering Pipeline 10 min



Practice Quiz: Practice Quiz (Ungraded) - ML Pipelines 2 questions



Quiz: SparkML concepts 3 questions

Due Sep 28, 1:59 AM CDT

Unsupervised Learning with Apache SparkML



Video: Introduction to Clustering: k-Means 3 min



Video: Using K-Means in Apache SparkML 2 min



Reading: Exercise 2 - Working with Clustering and Apache SparkML 10 min



Practice Quiz: Practice Quiz (Ungraded) - SparkML Algorithms 2 questions

