Machine Learning for Big Data

Introduction

Introductions

Robert Dempsey

- 18 years in tech
- Degrees in Comp Sci & MBA
- Founded 3 businesses
- Founder, Data Wranglers DC
- Books
  - Python Business Intelligence Cookbook (2015)
  - Building Machine Learning Pipelines (2018)
- International speaker
- Instructor & consultant at District Data Labs
- I drink a LOT of coffee :)

Gimme Your Infos!

1. Your name
2. What your role is
3. Why you are attending this course
4. What you hope to get out of this course

Course Agenda

Course Setup

- 8, half days
  - Interactive lecture + hands-on lab
- Yes, there will be breaks
- Everyone gets a number!

Day One: Data Analytics with Hadoop

- Introduction to Distributed Computing
  - The Age of Data Products
  - Building Data Products at Scale
  - Data Product Architectures
- Hadoop: An Operating System for Big Data
  - Hadoop Architecture
  - What's In A Cluster?
  - HDFS Caveats

Day Two: Data Analytics with Hadoop

- Setting Up for Big Data Analytics
  - Building a Hadoop Cluster in AWS
  - Building a Spark Cluster in AWS
  - Configuring Your Local Environment
- Introduction to Spark
  - Building Applications for Spark
  - Writing Spark Applications

Day Three: Machine Learning on Big Data

- Machine Learning Overview
- Model Categories & Types of Output
- Operationalizing Machine Learning
- Threats to Machine Learning

Day Four: Machine Learning on Big Data

- Big Data Approaches
- Sampling and Fitting in Memory
- A Tour of Model Families

Day Five: Supervised Machine Learning

- Overview of Supervised Learning
- Regression Models (Algorithms)
- Model Evaluation
- Hands-on Lab: Regression at Scale

Day Six: Supervised Machine Learning

- Classification Models (Algorithms)
- Model Evaluation

Day Seven: Unsupervised Machine Learning

- Overview of Unsupervised Learning
- Distance Metrics
- Clustering Algorithms

Day Eight: Unsupervised Machine Learning

- Clustering
  - Algorithms Review
  - Evaluation
  - Visualization
- Clustering at Scale

Building a Data Machine


Input
Output