Practical Machine Learning with R

Instructor

Matthew Renze
Twitter: @matthewrenze
Email: info@matthewrenze.com
Web: http://www.matthewrenze.com

Course Description

R is a very popular open-source programming language for machine learning. Its interactive programming environment and powerful data analysis capabilities make R an ideal tool for machine learning.

This workshop will provide an introduction to the R programming language using RStudio. In addition, we will demonstrate how we can use R to train a series of machine learning models. We'll cover supervised and unsupervised learning in the form of classification, regression, and clustering. We'll also learn how to handle Big Data and train deep neural networks. Finally, we'll learn how to deploy these models to production.

Prerequisites

Please bring your own Windows laptop and complete Lab 0 to install all of the necessary software before the workshop begins.

Module Descriptions

1. **Introduction** – introduce machine learning and the R programming language
2. **Classification** – learn how to predict categorical outcomes
3. **Regression** – learn how predict numeric outcomes
4. **Clustering** – learn how to predict groups of data based on similarity
5. **Model Tuning** – learn how to evaluate and optimize models
6. **Ensemble Learning** – learn how to train, test, and implement ensembles of models
7. **Deep Learning** – learn how to train, test, and implement deep learning models
8. **ML in Practice** – learn how to deploy machine learning solutions into production

Learning Objectives

When students are finished with this workshop, they should understand the following:

Introduction
- What machine learning is, why it is important, and how the machine learning process works
- What R is and why it has become so popular for machine learning
- How to create data types, data structures, subset data tables, and find help on R topics

Classification
- What classification is, how it works, and applications for classification
- What are decision trees and neural networks
- How to train, test, and make predictions with a classifier

Regression
- What regression is, how it works, and applications for regression
- What are linear regression, non-linear regression, and multiple regression
- How to train, test, and make predictions with regression

Clustering
- What clustering is, how it works, and applications for clustering
- What are k-Means and hierarchical clustering
- How to group similar data points using clustering algorithms

Model Tuning
- What are overfitting, underfitting, and regularization
- How to evaluate machine learning model performance
- How to tune hyperparameters to optimize machine learning models

Ensemble Learning
- What ensemble learning is, why it's important, and how it works
- What are random forests, boosting, ??? and how they work
- How to train, test, and make predictions with ensemble classifiers

Deep Learning
- What deep learning is, why it's important, and how it works
- What a deep neural network is and how it works
- How to train, test, and make predictions with a deep neural network

ML in Practice
- What are the options for deploying machine learning into production
- How to deploy machine learning models into production with R
- What industry best practices exist for using R for machine learning

Course Outline

Introduction

Lecture

- What is machine learning?
- Why is machine learning important?
- How does the machine learning process work?
- What is R?
- Why is R so popular for machine learning?
- R language basics

Lab

- Installation and setup
- Hello World
- Working with data types
- Working with data structures
- Working with data frames
- Miscellaneous topics

Classification

Lecture

- Classification
- Decision tree classifier
- Neural network classifier
- Training and test data sets

Lab

- Creating a training and test data set
- Predicting categories with a decision tree
- Predicting categories with a neural network

Regression

Lecture

- Regression
- Simple linear regression
- Non-linear regression
- Multiple variable regression

Lab

- Predicting values with simple linear regression
- Predicting values with a neural network

Clustering

Lecture
- Clustering
- k-Means clustering
- Hierarchical clustering

Lab
- Grouping data with k-means clustering
- Grouping data with hierarchical clustering

Model Tuning

Lecture
- Overfitting, underfitting, and regularization
- Evaluating model performance
- Tuning hyperparameters to optimize models

Lab
- Selecting a model based on accuracy, precision, and recall
- Tuning hyper-parameters for a model

Ensemble Learning

Lecture
- What is ensemble learning?
- Random forests, boosting, ???

Lab
- Classification with random forests
- Regression with random forests

Deep Learning

Lecture
- What is deep learning?
- Deep learning techniques
- Deep learning applications

Lab
- Making predictions with a deep neural network

Machine Learning in Practice

Lecture
- Deploying ML in production with R
- Best practices for ML

Lab
- Deploying a ML model to production with Shiny