Microsoft

The Microsoft R Laboratory

X – X April 2016
MTC, Location, Country

Summary

The Microsoft R Laboratory is a 2-day training course that explores the following topics through a series of hands-on labs using Microsoft R Server:

- Data Cleansing
- Feature Engineering
- Building regression, classification and clustering models
- Operationalizing R with
  - Azure Machine Learning
  - DeployR
- Real-time R analytics with:
  - Azure Stream Analytics
  - DeployR (RBrokerFramework)
- Power BI Dashboards
- Speeding up pleasingly parallel jobs (e.g. permutation feature selection, ensemble modelling, simulation, etc)

Please note: This is a hands-on training course. PowerPoint presentations are kept to a minimum (~2 hours over two days).

Audience

Given the technical nature of this training course the target audience are data scientists with **strong R skills**. There are two labs that require a basic knowledge of C#.

Prerequisites

- **An Azure subscription with admin rights (during the course participants will create a data science virtual machine, which will be shutdown/deleted at the end of each lab).** Participants should use their organisation's Azure enterprise agreement or have signed up to the one-month free trial (click here).

- Ideally a PowerBI account – one of the labs uses PowerBI to create a real-time dashboard. Should you not have a PowerBI account, we will offer an alternative solution.

Cost

The training is provided free of charge. However, **participants are expected to meet the costs of their own travel and expenses**.

Location

Microsoft Training Centre, Building 3, Microsoft Campus, Thames Valley Park, UK.

Start/Finish Times

Day 1: 09:30 – 17:15
Day 2: 09:00 – 17:15

Catering

Tea, coffees and lunch will be provided each day.

Dress code

Casual

Day 1 – Agenda

Time	Session	Detail
09:30-09:45	Coffee, introductions and getting logged on	
09:45-11:00	Presentation: Microsoft R Server	In this session we provide an in-depth presentation on the concepts of Microsoft R Server including: • How MRS builds on top of open source R • 'rx functions' • Distributed compute contexts (SQL Server, Hadoop, Spark and Teradata) • Data connectors (Text, XDF, SAS, SPSS, ODBC) • Deployment of R Scripts
11:00-12:00	Lab: Introduction to Microsoft R Server	This lab is intended to serve as an introduction to Microsoft R Server taking through data import, basic feature engineering and modelling. Upon completing this lab, you will have hands-on experience with the following functions and concepts related to Microsoft R Server: • E**X**ternal **D**ata **F**rame (XDF) file format: Performance gains, best practice, etc. • The rxDataStep function for data management • Exploring summary statistics about datasets and features • Training a Linear Regression model using rxLinMod • Testing a trained model with rxPredict • Evaluating model performance with rxRoc
12:00-13:00	Lunch	
13:00-14:30	Lab: Data Cleansing and Management with Microsoft R Server	This lab explores leveraging the best of SQL Server and Microsoft R Server to cleanse, preprocess and uncover predictively useful features. Upon completing this lab, you will have hands-on experience with the following functions and concepts related to Microsoft R Server: • Running SQL queries in-database through RxOdbcData • Identifying and removing predictors with missing values • Identifying and removing zero variance predictors • Identifying and removing highly correlated predictors • Identifying and removing linear dependencies • Feature selection using mutual information • Feature selection using Decision forests
14:30-15:30	Lab: Building Predictive models with Microsoft R Server	This lab explores how to build and validate predictive models on large datasets using Microsoft R Server. Upon completing this lab, you will have hands-on experience with the following functions and concepts related to Microsoft R Server: • Splitting data into training and test sets with rxSplit • Building a logistic regression model • Building a decision forest model • Building a K-Means model • Building a Naïve Bayes model • Comparing models on unseen data using rxPredict and rxRoC / rxGini
15:30-15:45	Break	
15:45-17:00	Free Lab: Microsoft R Server	In this lab you can choose a "recipe" to further explore Microsoft R Server. However, the recipes are light on detail to aid your exploration. You can choose from the following recipes: • Any Data! • Clustering single malt whiskies
17:00-17:15	Wrap up	In this session we will wrap-up the day with a quick Q&A

Day 2 – Agenda

Time	Session	Detail
09:00-09:30	Presentation: R Deployment options	In this session, we overview the deployment (web services) options for R-based analytics in the cloud (Azure Machine Learning) and also on-premise/cloud DeployR.
09:30-11:00	Lab: Operationalizing R with Azure Machine Learning	In this lab we will explore the functionality of the AzureML R package. Upon completing this lab, you will have hands-on experience with the following functions and concepts related to operationalizing R with Azure Machine Learning: • Workspace: connect to and manage AzureML workspaces • Datasets: upload and download datasets to and from AzureML workspaces • Publish: define a custom function or train a model and publish it as an Azure Web Service • Consume: use available web services from R in a variety of convenient formats
11:00-12:30	Lab: Getting started with DeployR	In this lab you will create some applications that consume R as a web service. Upon completing this lab, you will have hands-on experience with the following functions and concepts related to DeployR: • How to connect a client application to DeployR • Authenticating and versioning in DeployR • Passing objects to/from an R session • Uploading data files from a client to the DeployR server for processing • Interactive plots with DeployR • How to integrate DeployR into Excel • Real-time analytics with DeployR
12:30-13:30	Lunch	Prior to Lunch we will provision an R Server on Spark cluster for use in the afternoon session.
13:30-14:00	Presentation: Introduction to MRS on Hadoop	In this session we provide an overview of the features of Microsoft R Server on Hadoop.
14:00-16:00	Lab: Getting Started with Microsoft R Server on HDInsight (Spark)	This lab shows how to get started using Microsoft R Server on Hadoop using the publically available “airlines” dataset. Upon completing this lab, you will have hands-on experience with the following concepts: • Provisioning an R Server on Spark cluster • Cleaning and joining datasets using SparkR in R Server • Train, score and evaluate using R Server on Spark • Publish model as a web service from R to Azure Machine Learning
16:00-16:30	Wrap-up	In this session we will wrap-up the day with a quick Q&A