Microsoft

Microsoft R Server Free Lab

Contents

Overview

Prerequisites

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. Therefore, before embarking on this lab it is a good idea to have completed the following labs:

- Introduction to Microsoft R Server
- Data Cleansing & Management with Microsoft R Server
- Building Predictive models with Microsoft R Server

Option 1 – Any Data!

Choose any dataset you like either:

- In the SampleData directory provided with Microsoft R Server i.e. list.files(rxOptions()$sampleData)
- Use a publically available dataset – https://www.kaggle.com/datasets – contains a list that you can use.

Use Microsoft R Server to:

1. Convert data to XDF file format
2. Cleanse data
3. Feature engineering
4. Split the data into training/test/validation sets
5. Normalise data (if required)
6. Model Data / find insight in the data

How would you operationalize both the scoring and re-training of the model (model life cycle)?

Option 2 – Clustering Single Malt Whiskies

In the folder where this document resides is a file called whiskies.txt. This is a small dataset containing 86 records where each row represents a single malt whisky distillery in Scotland and the columns are flavour scores (1-5) given for on Body, Honey, Spicy, Malty, etc.

A blog has been written on using Open Source R to segment (based on kmeans clustering) the distilleries based on their flavours:

http://www.r-bloggers.com/k-means-clustering-86-single-malt-scotch-whiskies/

How would you make this solution scale using ScaleR? In particular, you will need to the following analytics to work on big data:

- An rxDataStep to scale and center each flavor column
- Produce a sum within squares error function to find the number of clusters to use in the dataset (**hint:** this function will use rxSummary and rxKmeans).

- Run an rxKmeans to segment the whiskies into different clusters.

How would you operationalize both the scoring and re-training of the model (model life cycle)?

Terms of Use

By using this Hands-on Lab, you agree to the following terms:

The technology/functionality described in this Hands-on Lab is provided by Microsoft Corporation in a “sandbox” testing environment for purposes of obtaining your feedback and to provide you with a learning experience. You may only use the Hands-on Lab to evaluate such technology features and functionality and provide feedback to Microsoft. You may not use it for any other purpose. You may not modify copy, distribute, transmit, display, perform, reproduce, publish, license, create derivative works from, transfer, or sell this Hands-on Lab or any portion thereof.

COPYING OR REPRODUCTION OF THE HANDS-ON LAB (OR ANY PORTION OF IT) TO ANY OTHER SERVER OR LOCATION FOR FURTHER REPRODUCTION OR REDISTRIBUTION IS EXPRESSLY PROHIBITED.

THIS HANDS-ON LAB PROVIDES CERTAIN SOFTWARE TECHNOLOGY/PRODUCT FEATURES AND FUNCTIONALITY, INCLUDING POTENTIAL NEW FEATURES AND CONCEPTS, IN A SIMULATED ENVIRONMENT WITHOUT COMPLEX SET-UP OR INSTALLATION FOR THE PURPOSE DESCRIBED ABOVE. THE TECHNOLOGY/CONCEPTS REPRESENTED IN THIS HANDS-ON LAB MAY NOT REPRESENT FULL FEATURE FUNCTIONALITY AND MAY NOT WORK THE WAY A FINAL VERSION MAY WORK. WE ALSO MAY NOT RELEASE A FINAL VERSION OF SUCH FEATURES OR CONCEPTS. YOUR EXPERIENCE WITH USING SUCH FEATURES AND FUNCITONALITY IN A PHYSICAL ENVIRONMENT MAY ALSO BE DIFFERENT.

FEEDBACK. If you give feedback about the technology features, functionality and/or concepts described in this Hands-on Lab to Microsoft, you give to Microsoft, without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft software or service that includes the feedback. You will not give feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your feedback in them. These rights survive this agreement.

MICROSOFT CORPORATION HEREBY DISCLAIMS ALL WARRANTIES AND CONDITIONS WITH REGARD TO THE HANDS-ON LAB, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. MICROSOFT DOES NOT MAKE ANY ASSURANCES OR REPRESENTATIONS WITH REGARD TO THE ACCURACY OF THE RESULTS, OUTPUT THAT DERIVES FROM USE OF THE VIRTUAL LAB, OR SUITABILITY OF THE INFORMATION CONTAINED IN THE VIRTUAL LAB FOR ANY PURPOSE.

DISCLAIMER

This lab contains only a portion of the features and enhancements in Microsoft Azure. Some of the features might change in future releases of the product.