

# **Principles of Machine Learning**

Lab Setup Guide

#### Overview

This guide takes you through the steps to create an environment for performing the data science experiments described in this repository.

To prepare the lab environment, you must perform the following tasks:

- 1. Create an Azure ML account
- 2. Download and extract the lab files
- 3. Optionally, install Microsoft R Open and RStudio or Python Anaconda

#### What You'll Need

To perform the setup tasks, you will need the following:

- A Windows, Linux, or Mac OSX computer.
- A web browser and Internet connection.

# Create an Azure ML Account

Azure ML offers a free-tier account, which you can use to complete the labs in this repository. **Note**: A free Azure ML workspace is <u>not</u> the same as a Microsoft Azure trial subscription!

#### Sign Up for a Microsoft Account and a Free Azure ML Workspace

- 1. If you do not already have a Microsoft account, sign up for one at <u>https://signup.live.com/</u>.
- If you do not already have an Azure ML workspace, browse to <a href="https://aka.ms/edx-dat203.2x-aml">https://aka.ms/edx-dat203.2x-aml</a> and click **Get Started Now**. Then follow the instructions to sign up for a free Azure ML workspace. If prompted, sign in with your Microsoft account credentials. **Note**: Your free-tier Azure ML workspace allows you unlimited access, with some reduced capabilities compared to a full Microsoft Azure subscription. Your experiments will only run at low priority on a single processor core. As a result, you will experience some longer wait times. However, you have full access to all features of Azure ML.

# Download the Lab Files

All code and data files you will need are contained in a zip file. Follow these steps to download and install the lab files.

- 1. Download the lab files from <a href="https://aka.ms/edx-dat203.2x-labfiles">https://aka.ms/edx-dat203.2x-labfiles</a>.
- 2. Extract the downloaded zip file to a convenient folder on your local computer.

# Optional: Install Microsoft R Open and RStudio

R is a programming language for conducting statistical analysis processes and visualizing data. You will use R code in Azure ML for some of the labs in this course, and having a fully-featured R editor can be useful if you want to explore this code more thoroughly.

**Note**: For this course, you can choose to complete programming exercises in Python or R (or both if you are truly ambitious!). If you plan to use R, complete this procedure to install the R runtime and development tools. If you do not plan to use R, you can skip this section and go to the procedure for installing Anaconda Python.

#### Install Microsoft R Open

**Note**: Microsoft R Open is an enhanced version of R. Microsoft R Open uses the multi-core MKL math libraries to enhance performance. However, if are using GNU R downloaded from CRAN, you can continue to do so for this course.

- 1. In a web browser, navigate to https://aka.ms/edx-dat203.2x-r.
- 2. Select the **Microsoft R Open** download for your operating system and following the installation directions.

**Note**: If you are using Mac OS X, you can skip the following step as the MKL math libraries are already included.

- 3. When the Microsoft R Open installation has finished, select the **MKL** math library download for your operating system and follow the installation directions.
- Verify your installation by starting Microsoft R Open and entering a simple R expression such as 1
   + 2 (which should produce the result 3) in the console as shown in the following image.

```
- C ×
R Console
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86 64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
  Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Microsoft R Open 3.2.3
Default CRAN mirror snapshot taken on 2016-01-01
The enhanced R distribution from Microsoft
Visit http://go.microsoft.com/fwlink/?LinkID=722555 for information
about additional features.
Multithreaded BLAS/LAPACK libraries detected. Using 4 cores for math algorithms.
[Previously saved workspace restored]
> 1 + 2
[1] 3
>
```

5. Close R.

#### Install RStudio

- 1. In a web browser, navigate to https://www.rstudio.com/products/rstudio/download/.
- 2. Run the installer for your operating system (Windows, MacOSX, Ubuntu, or Fedora) to install RStudio.
- 3. To verify installation, start RStudio, and look at the console which should resemble the following image:

```
Console C:/Users/Steve/Dropbox/Azure ML/Data Science Essentials/Mod 4/ 🔗
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Microsoft R Open 3.2.3
Default CRAN mirror snapshot taken on 2016-01-01
The enhanced R distribution from Microsoft
Visit http://go.microsoft.com/fwlink/?LinkID=722555 for information
about additional features.
Multithreaded BLAS/LAPACK libraries detected. Using 4 cores for math algorithms.
[Workspace loaded from C:/Users/Steve/Dropbox/Azure ML/Data Science Essentials/Mod 4/.RDa
al
> 2 + 3
[1] 5
> |
```

-

- 4. Verify that RStudio is configured to use the current version Microsoft R Open, by noting the version of Microsoft R Open displayed on the console. (The current version is 3.2.3).
- 5. If your configuration is not correct select **Global Options** on the **Tools** menu and set the path to the directory where you installed Microsoft R Open, as shown in the following image.

Console C:/Users/Steve/Dro	opbox/Azure ML/Dat	a Science Essentials/Mod 4/ 🔗	- [		
R version 3.2.3 (2015-12-10) "Wooden Christmas-Tree" Copyright (C) Options Platform: x86					
R is free soft You are welcom Type 'license( R is a collabo	General	R version: [Default] [64-bit] C:\Program Files\Microsoft\MRO\R-3.2.3 Change Default working directory (when not in a project): ~ Browse			
Type 'contribu 'citation()' o	C Appearance	<ul> <li>Restore most recently opened project at startup</li> <li>Restore previously open source documents at startup</li> </ul>			
Type 'demo()' 'help.start()' Type 'q()' to	Pane Layout	Restore .RData into workspace at startup Save workspace to .RData on exit: Ask			
Microsoft R Op Default CRAN m The enhanced R Visit http://g about addition	Packages	<ul> <li>Always save history (even when not saving .RData)</li> <li>Remove duplicate entries in history</li> <li>Use debug error handler only when my code contains errors</li> <li>Automatically expand tracebacks in error inspector</li> </ul>			
Multithreaded [Workspace loa a]	Spelling Git/SVN	Default text encoding:         [Ask]         Change         Automatically notify me of updates to RStudio	<b>jorithms.</b> ≥ntials/Mod 4/.RDa		
> 2 + 3 [1] 5 >		OK Cancel Apply			

6. When you have verified installation and configuration, close RStudio.

#### Install R packages

Several of the labs for this course require R packages that are not installed by default. These packages include:

- ggplot2
- gridExtra
- dplyr

Follow these instructions to install these packages:

- 1. In RStudio, locate the pane with **Packages** tab and click on it.
- 2. Click the Install icon.
- 3. In the Packages text box of the dialog, type '**ggplot2**, **dplyr**, **gridExtra**', ensuring the names are comma separated and the **Install dependencies** box is checked, as shown in the figure.

Files	Plots	Packages	Help	Viewer			
0.1	nstall	🖸 Update	📕 Pac	krat 🕓		Q	L
_	Name				Description	Version	
User	Library						
	assertth	at	1	Install Packa	ges	0.1	8
	base64	enc		Install fron	Configuring Repositories	0.1-2	8
	BH			Repositor	y (CRAN) 🔻	1.58.0-1	
	bitops			Packages	separate multiple with space or comma):	1.0-6	0
	brew			ggplot2, d	Iplyr, gridExtra	1.0-6	8
	caTools				·······	1.17.1	0
	chron			Install to L	brary:	2.3-46	۲
	colorsp	ace		C:/Users/:	steve/Documents/R/win-library/3.2 [Default]	1.2-6	۲
	curl			✓ Install of	lependencies	0.9	0
	data.tal	le				1.9.4	0
	datadr				Install Cancel	0.8.1	8
	DBI		۰.			0.3.1	0
	devtool	5			Tools to Make Developing R Packages Easier	1.10.0	0

- 4. Click **Install**. Expect a great deal of text to appear on the console. Watch for error messages. Note that you may be prompted to create a user library for your downloaded packages.
- 5. After the packages have been installed, at the console prompt in RStudio, type the following commands to test loading the packages:

```
library(ggplot2)
library(dplyr)
library(gridExtra)
```

# **Optional: Install Python Anaconda**

You will use Python code in Azure ML for some of the labs in this course, and having a fully-featured Python editor can be useful if you want to explore this code more thoroughly.

**Note**: In this repository, you can choose to complete programming exercises in Python or R (or both). If you plan to use Python, complete this procedure to install the Python runtime and development tools. If you do not plan to use Python, you can skip this procedure.

#### Install the Python Anaconda Distribution

**Note**: Microsoft Azure Machine Learning uses a version of the Anaconda Python distribution. However, if you are used to using another scientific Python distribution, such as Canopy, it is likely that you will be able to perform all of the exercises for this course using that distribution.

- 1. In a web browser, navigate to http://continuum.io/downloads.
- 2. Choose the installer for your operating system (Windows, Apple Macintosh, or Linux).
- 3. Complete the installation process for Python 2.7 (Not Python 3.X).
- 4. After installation is complete, verify the installation by starting Spyder, which should look similar to the following image:



#### Verify Connection to the IPython kernel

In some cases, when you first start Spyder, you may find that your IPython console does not respond to the commands you type at the prompt. If this occurs, wait a while to allow IPython kernel to start.

If the IPython console continues to be unresponsive, you may need to manually connect to an IPython kernel manually. Follow the steps below to connect a console to an IPython kernel.

1. From the kernel icon in the upper right above the IPython console window, select **Connect to an existing kernel** as shown in the figure below:

				X
		<b></b>	÷	+
				₽×
SC v.	1 😔	Restart kernel	Ctr	'l+.
	IP:	Open a new console	Ctr	1+T
		Connect to an existing ke	rnel	
	_			

2. Select **Browse** on the dialog and then **Open** the first kernel on the list shown in the figure below:

Please enter the connection info you can either select its JSON write directly its id, in kernel-3764.json or just 3 Connection info: Path to conne	o of the kernel you want to connect to. For that connection file using the Browse button, or case it's a local kernel (for example 764). ction file or kernel id Browse			
<ul> <li>Open IPython connection</li> <li>Steve </li> </ul>	file .ipython > profile_default > security		<b>- ←</b> Sec	arch security
Organize 🔹 New fold	er			:= • 🗌 🤅
★ Favorites	Name	Date modified	Туре	Size
E Desktop	🞜 kernel-2516	4/11/2016 6:41 AM	JSON File	=
😹 Recent Places 😑	kernel-606 Type: JSON File	31/2016 10:32 A	JSON File	
iCloud Photos	kernel-957 Size: 258 bytes	3/2016 7:21 PM	JSON File	
🐌 Downloads	5 kernel-2d6, Date modified: 4/11/2016 6:41 A	M /24/2015 4:42 PM	JSON File	
😌 Dropbox	🗊 nbserver-6164	11/24/2015 4:39 PM	JSON File	No prev availab
	√ kernel-4bd4cb8c-7701-45f5-ad19-ba997	11/24/2015 7:56 A	JSON File	avanab
🚞 Libraries	🗊 nbserver-7228	11/24/2015 7:55 A	JSON File	
Documents	<b>IJ</b> kernel-4ff50dd4-7cbe-4fc4-87c6-c64608	10/27/2015 12:30	JSON File	
🕹 Music	🗊 nbserver-8828	10/27/2015 12:30	JSON File	
Pictures		<u>10/22/2015 Φ.49 Δ</u>	ISON File	
File na	me: kernel-2516		▼ *.json	-
			Open	Cancel

- 3. Click OK, to complete selection of the Python kernel for your IPython session.
- 4. In the IPython Console pane, enter a command to verify that everything is working for example, enter **1**+ **2** and verify that the answer **3** is returned.
- 5. Close Spyder.

### Summary

By completing the tasks in this setup guide, you have prepared you environment for the labs in this course. Now you're ready to start learning how to build data science and machine learning solutions.