

ML Workflow for Research Scientists

Ajinkya K Mulay

2021-02-06

Contents

1	Preamble	5
	Who is this wiki for?	5
	What won't you learn?	5
	What will you learn?	5
2	Python and Google Colaboratory	7
3	Git, Github and Desktop management	9
4	Python and Google Colab Setup	11
	Importing Datasets	11
	Importing Local Files	11
	4.1 Importing models between multiple projects	11
5	Unit Testing	13
	5.1 Example one	13
	5.2 Example two	13
6	Documentation	15
7	Putting It All Together	17
	7.1 R Markdown	17
	7.2 Including Plots	17

Chapter 1

Preamble

Who is this wiki for?

There is a long list of Python and Github workflow tutorials. There are so many Machine Learning Tutorials too. However, there is no one easy to follow Machine Learning-ML Cloud Service-Local Development-Benchmarking Workflow. I am going to talk about three Free services which any person can use with basic internet access - Google Colaboratory (for ML Cloud Service), VS Code (Local ML Development) and Weights and Biases (for benchmarking). TL;DR- This is a process guide for deploying your ML models as a researcher!

What won't you learn?

How to code in Python or develop Machine Learning Models or learn Pytorch.

What will you learn?

- How to contribute code to a well-structured projects
- Work with multiple Google Colaboratory Notebooks
- Create easy Python Documentation
- Scale up as your project and team grows in complexity and size
- Benchmark and Project Tracking

Chapter 2

Python and Google Colaboratory

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(  
  head(iris, 20), caption = 'Here is a nice table!',  
  booktabs = TRUE  
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).



Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 3

Git, Github and Desktop management

Here is a review of existing methods.

Chapter 4

Python and Google Colab Setup

Importing Datasets

Importing Local Files

4.1 Importing modelos between multiple projects

Chapter 5

Unit Testing

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Documentation

We have finished a nice book.

Chapter 7

Putting It All Together

7.1 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

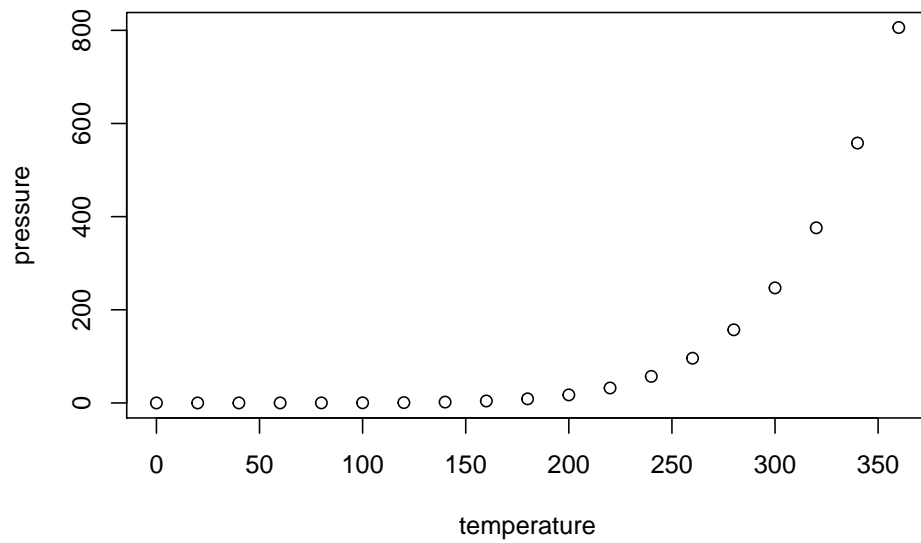
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.:26.00
## Median :15.0    Median :36.00
## Mean   :15.4    Mean   :42.98
## 3rd Qu.:19.0    3rd Qu.:56.00
## Max.   :25.0    Max.   :120.00
```

7.2 Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.21.