

Google Colab Research

- Google cloud in the form of virtual laboratory for machine learning research
- The user must have a google account (which will automatically own a cloud file storage as “Google Drive”)
- Programming is done in the “notebook” style
- Most essential libraries are already installed, ready for import as module
- Using Colab automatically create folder called “Colab Notebooks” in Google Drive
 - Your programs should be kept in this folder
- Using Colab frees your PC from handling the data mining load
- Colab cloud has a large number of GPUs which helps speed up the learning process

- 1) With your google account signed-in, go to <https://colab.research.google.com>
- 2) Create a new notebook. The notebook will be created as an “untitled” file with .ipynb extension
- 3) Try a simple Python 3 code. Then click the play button to run the cell.
- 4) Change the file name. There are more than one way to do this. Choose whatever you prefer.
- 5) In order to access a file in Google Drive/Colab Notebooks, your google drive must be mounted to the Colab.

```
from google.colab import drive
drive.mount('/content/gdrive')
%cd /content/gdrive/My\ Drive/Colab\ Notebooks
```

Running this cell will prompt you to sign in to google account and obtain the authorization code. Once provided with the authorization code, the drive is mounted. Then, the “changing directory” command (cd) will switch the working directory to “Colab Notebooks”.

- 6) Create a simple text file in your “Colab Notebooks” folder. And use the following code to test if your notebook can read a text file in Google Drive.

```
infile = open('testfile.txt', 'r')
for line in infile:
    print(line.strip())    #each line ends with end-of-line character
```

Numpy

- Numpy is a well-known library that provides support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- Machine learning libraries integrate well with Numpy arrays

1) Test importing numpy module into your Colab notebook.

If you use Jupyter lab, you probably need to separately install Numpy. Installation varies from OS to another, so study for yourself on how to install.

2) Test use Numpy with the following simple code.

```
import numpy as np

a = np.array([[1,2,3],[4,5,6]])
print(a)
print(a.shape)
```

3) It is not suggested to spend time learning all features of Numpy. However, [NumPy Getting Started \(w3schools.com\)](https://www.w3schools.com/numpy/) is suggested for learning conventions and operations for Numpy's multi-dimensional arrays.

Make sure that you can learn the necessary features very fast when you need to.

Pandas

- Pandas is a Python library used for working with data sets.
- It has functions for analyzing, cleaning, exploring, and manipulating data.
- Pandas analysis is mainly statistical.
- Pandas' "Data Frames" is effective for keeping data sets in your program. However, using both Pandas data frames and Numpy array can be confusing as the data frames can manage heterogenous data types while the Numpy array mandates a homogenous data type. Acquiring skills with either is left to your discretion. But not being skillful enough to use any will trouble you throughout the course.

- 1) Try importing Pandas module into your Colab notebook.

If you use Jupyter lab, you probably need to separately install Pandas. Installation varies from OS to another, so study for yourself on how to install.

- 2) Test use Pandas with the following simple code.

```
import pandas as pd

data = {
    "calories": [420, 380, 390],
    "duration": [50, 40, 45]
}

#load data into a DataFrame object:
df = pd.DataFrame(data)

print(df)
```

- 3) Just like Numpy, it is not suggested to spend time learning all features of Pandas. However, [Pandas Getting Started \(w3schools.com\)](https://www.w3schools.com/pandas/getting-started.asp) is suggested for learning conventions and operations for Pandas' data frames.

Make sure that you can learn the necessary features very fast when you need to.

Matplotlib

- Matplotlib is a low level graph plotting library in python that serves as a visualization utility.
- It is a useful tool to depict the data graphically, particularly for viewing characteristics of data sets or results.

- 1) Try importing matplotlib module into your Colab notebook.

If you use Jupyter lab, you probably need to separately install Matplotlib. Installation varies from OS to another, so study for yourself on how to install.

- 2) Test use Matplotlib with the following simple code.

```
import matplotlib.pyplot as plt
import numpy as np

xpoints = np.array([0, 6])
ypoints = np.array([0, 250])

plt.plot(xpoints, ypoints)
plt.show()
```

- 3) Unlike Numpy and Pandas, Matplotlib does not have too many features to learn. However, its usefulness depends on imagination which is in turn dependent on skills. Thus, going through a short tutorial of Matplotlib starting at [Matplotlib Getting Started \(w3schools.com\)](https://www.w3schools.com/matplotlib/matplotlib_getting_started.asp) is recommended.