

Practical Machine Learning



Practical Machine Learning

Lecture: Getting Started with Google Colab

Ted Scully



Getting Started with Google Colab

Google Colab is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

- Colab provides users with a free GPU (Tesla K80 GPU) or TPU. Currently the only restriction on usage is that there is a maximum duration of connection of a single VM instance of 12 hours.
- It comes preinstalled with essential packages such as NumPy, Pandas, Scikit-Learn, TensorFlow. You can also install additional software if needed.
- It is built on top of Jupyter Notebooks and presents a similar interface consisting of cells.
- Over the next few slides I will illustrate how to create Colab Notebook from your Google Drive, upload a dataset that resides in your drive and build a Keras Tensorflow model.
- First navigate to your Google Drive using your browser. You can directly create Colab Notebooks from there.



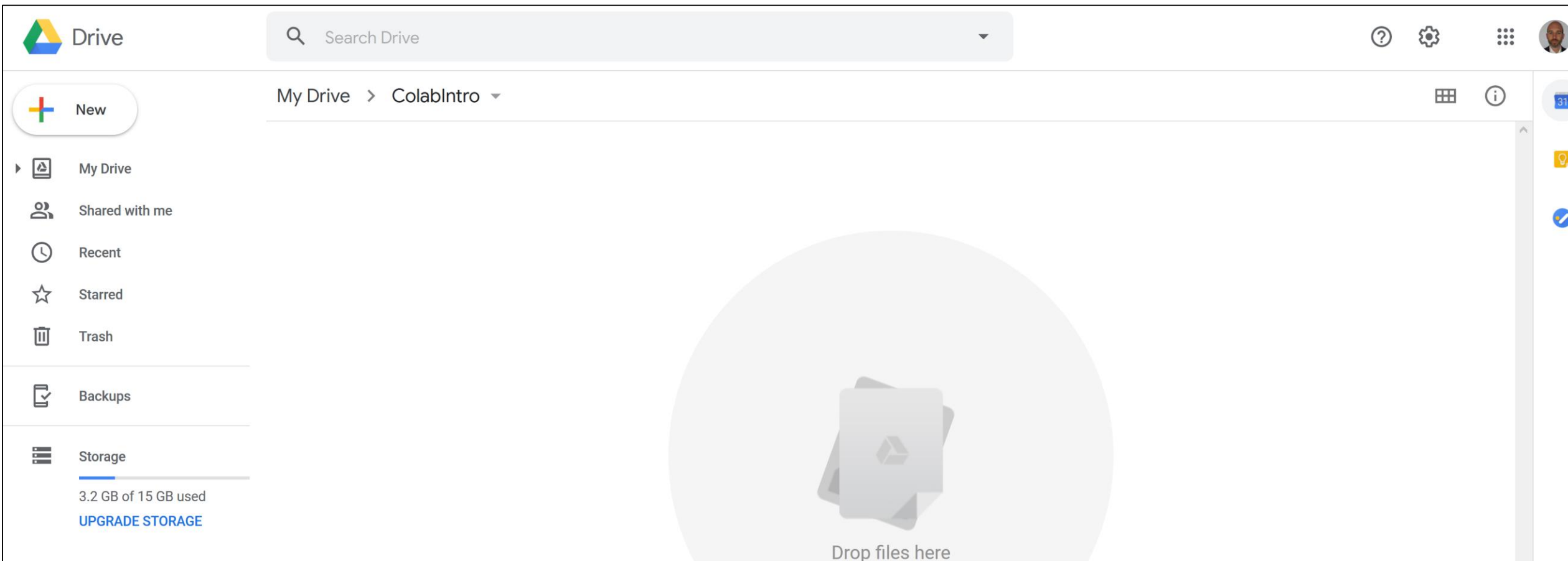
Getting Started with Google Colab

- In this short tutorial I will illustrate the following:
- Step 1. Create a Colab Notebook from your Google Drive
- Step 2: Mount a Google Drive dataset
- Step 3: Load the dataset using NumPy and perform a basic pre-processing step.

Please note you will need a Google (Gmail) account.

If you haven't used Google Drive before you can access it [here](#).

This image shows an empty folder that I have created in my Google Drive (the folder name is ColabIntro).



Add a new folder called small_mnist to your Google drive folder. This folder will contain a training set called small_mnist.csv
(You can find the small mnist folder and csv file on Canvas Week 1 unit. Just unzip the folder and drag to Google Drive)

The screenshot shows the Google Drive web interface. The address bar displays the URL: <https://drive.google.com/drive/u/0/folders/1kXiwlbXEAOt3DoeZqO5IrNP3qqXPr7In>. The left sidebar contains navigation options: 'New', 'My Drive', 'Shared with me', 'Recent', 'Starred', and 'Trash'. The main content area shows the breadcrumb 'My Drive > ColabIntro' and a table of files. The table has columns for 'Name', 'Owner', 'Last modified', and 'File size'. A single folder named 'small_mnist' is listed, owned by 'me', last modified at '11:42 AM me', with a file size of '-'. The folder is highlighted with a blue border.

Name	Owner	Last modified	File size
small_mnist	me	11:42 AM me	-



Drive



Search Drive



New



My Drive



Shared with me



Recent



Starred



Trash



Backups



Storage

3.2 GB of 15 GB used

[UPGRADE STORAGE](#)

My Drive > ColabIntro > small_mnist ▾



Name ↑

Owner

Last modified

File size





small_mnist.csv


me


11:42 AM me


17 MB


 Drive


 New


▶  My Drive


 Shared with me

 Recent

 Starred

 Trash

 Backups

 Storage

3.2 GB of 15 GB used
[UPGRADE STORAGE](#)

Search Drive


My Drive > ColabIntro ▾


Name ↑


Owner


 small_mnist


me


 New folder


 Upload files


 Upload folder


 Google Docs >


 Google Sheets >


 Google Slides >


 More >


 Google Forms >

 Google Drawings

 Google My Maps

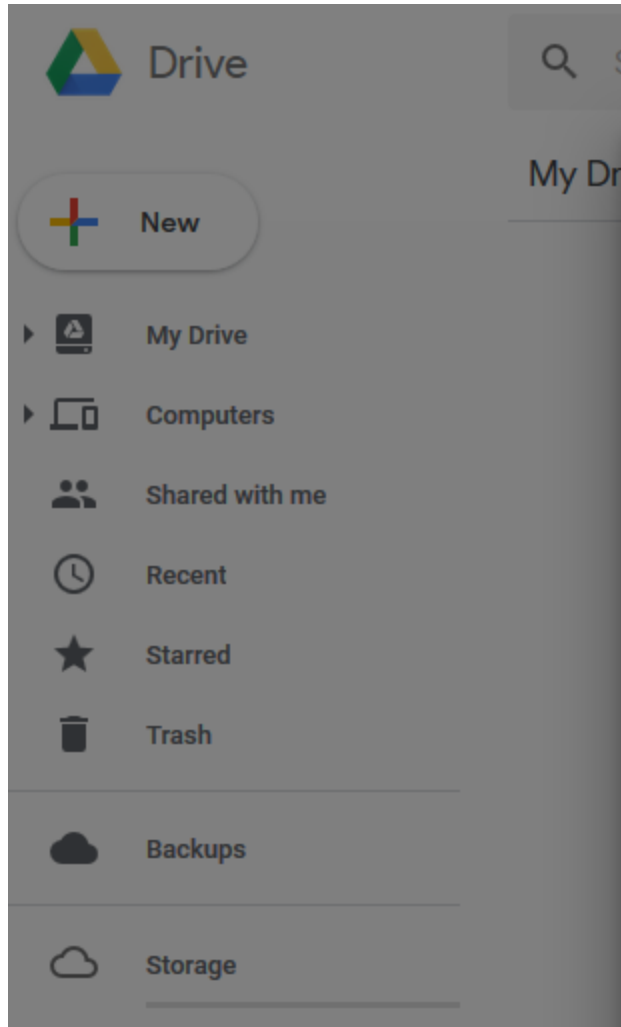
 Google Sites

 Google Jamboard

 Connect more apps

To create Colab notebooks we must add it as an app first as shown.

1. Right click and the dropdown menu shown will appear.
2. Click “connect more apps” and search for Colab.



Search Drive

My Drive

Connect apps to Drive

All ▾

colab



Colaboratory

offered by <https://colab.research.google.com>


A data analysis tool that combines code, output, and descriptive text into one collaborative document.






+ CONNECT

Productivity


★★★★★ (318)

Once you find Colab below click the connect button.

 New

-  My Drive
-  Shared with me
-  Recent
-  Starred
-  Trash


 Backups

 Storage


3.2 GB of 15 GB used


[UPGRADE STORAGE](#)


My Drive > ColabIntro ▾


Name ↑	Owner	
 small_mnist	me	11:42 AM me


The next time you right click you can now directly create a Colab notebook from this Google Drive folder.


 New folder

 Upload files


 Upload folder


 Google Docs >


 Google Sheets >


 Google Slides >


More >


 Google Forms >


 Google Drawings

 Google My Maps

 Google Sites

 Google Colaboratory

 Google Jamboard

 Connect more apps

The new Google Colab notebook opens automatically.

The notebook is given a default name (in this case Untitled0.ipynb). You can change this name by clicking on the title.





Search Drive

If you go back to your Google Drive you will now see the dataset folder and the Colab notebook.



New



My Drive



Shared with me





Recent



Starred

My Drive > ColabIntro

Name	Owner	Last modified
 small_mnist	me	11:42 AM me
 Untitled0.ipynb	me	11:55 AM me

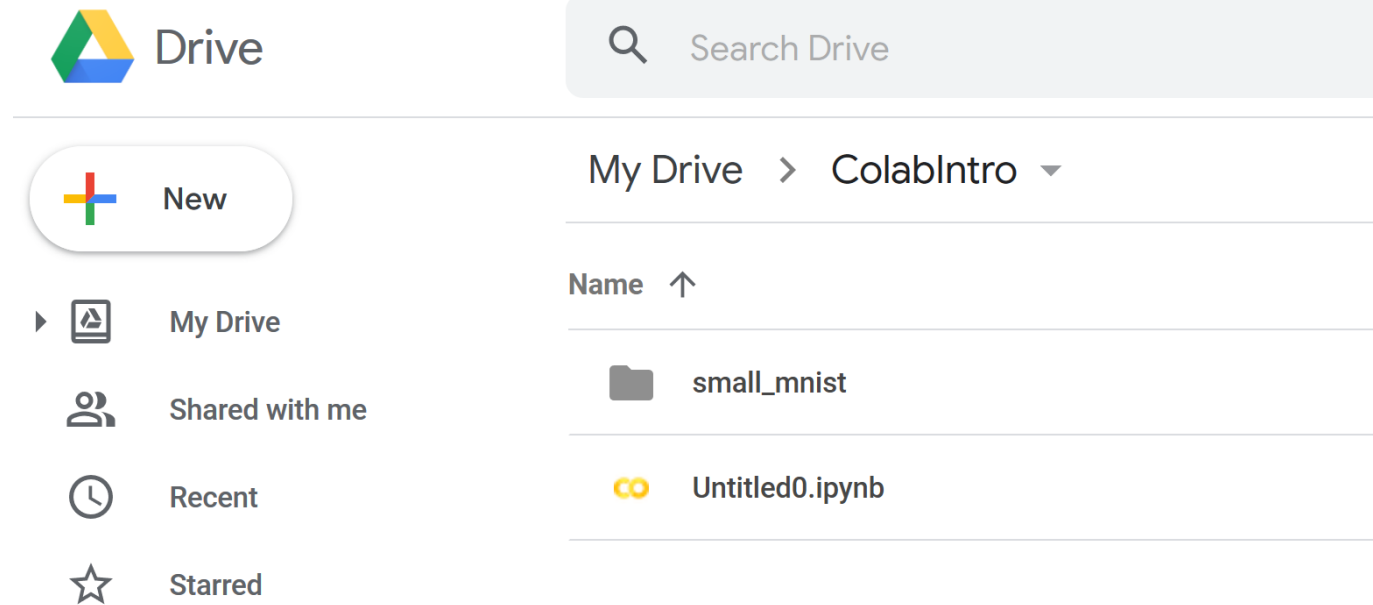
Components of Colab

- Colab is organized in a similar way to Jupyter notebooks. We can insert code or text cells in your notebook.
- Useful shortcuts include:
 - Running the currently selected cell is CTRL+Enter
 - CTRL+F9 will run all cells in the notebook
- We can also basic linux command line instructions such as pwd, ls, cd but putting a ! before the command
!ls

The next step is to mount the small mnist csv data file from Colab.

Mounting Google Drive using Colab

- Over the next few slides we will cover how to mount your Google drive and directly access data files on your drive from Colab.
- We will use the small MNIST dataset that I have stored in my Google Drive.
- You can mount your Google Drive in your Colab virtual machine using an authorization code. Once executed, you should then be able to access the data files on your Google Drive.
- The full Colab code for the following example can be obtained [here](#).



Mounting Google Drive

- Enter the following code into your Colab notebook to mount your Google Drive.

```
from google.colab import drive  
drive.mount('/content/gdrive')
```

- It will ask you to authenticate (click on the URL), once you do you should see a message as below that you drive has been mounted. You now be able to access your file



The screenshot shows the Google Colab interface for a notebook titled "Untitled0.ipynb". The top bar includes the Colab logo, a star icon, and a menu with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". On the right, there are buttons for "Comment", "Share", and a user profile icon. Below the top bar, there are tabs for "+ Code" and "+ Text". A status bar on the right shows "RAM" and "Disk" usage with progress bars, a green checkmark, and a button labeled "Editing". The main code area displays the following output:

```
[1] from google.colab import drive  
drive.mount('/content/gdrive')
```

Below the code, there is a message with a link icon: "Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.c". Below the link, it says "Enter your authorization code:" followed by a series of dots. At the bottom, it says "Mounted at /content/gdrive". A horizontal scrollbar is visible at the bottom of the code area.

Mounting Google Drive and Reading Data

- Now that the drive is mounted we can read in our data file using the following code (notice NumPy is already installed).
- Notice I use NumPy below to open this file in Colab.

```
import numpy as np

data = np.genfromtxt("/content/gdrive/My Drive/ColabIntro/small_mnist/small_mnist.csv", delimiter=",", skip_header=1)
```

Mounting Google Drive and Reading Data

- The labels are the first column in the CSV file so in the code below we separate the features (all columns except the first) from all other data.
- We then normalize the feature data. If you haven't heard of normalization, don't worry. We will be covering this later in the module.

```
# print the shape of the data
print (data.shape)

# extract the feature training data
dataFeatures = data[:, 1:]
print (dataFeatures.shape)

# normalize the data
dataFeatures = dataFeatures/255.0
```




+ Code + Text



```
[1] from google.colab import drive  
drive.mount('/content/gdrive')
```

Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g

Enter your authorization code:
.....
Mounted at /content/gdrive



```
[6] import numpy as np  
data = np.genfromtxt("/content/gdrive/My Drive/ColabIntro/small_mnist/small_mnist.csv", delimiter=",", skip_header=1)
```



```
# print the shape of the data  
print (data.shape)  
  
# pull out the feature training data  
dataFeatures = data[:, 1:]  
print (dataFeatures.shape)  
  
# normalize the data  
dataFeatures = dataFeatures/255.0
```



```
(10000, 785)  
(10000, 784)
```