Integrating Machine Learning into Geographic Research

Introduction to Google Colab

Developing machine learning models

- A machine learning model is a computer program; Like other computer programs, we need an integrated development environment (IDE) to develop machine learning model

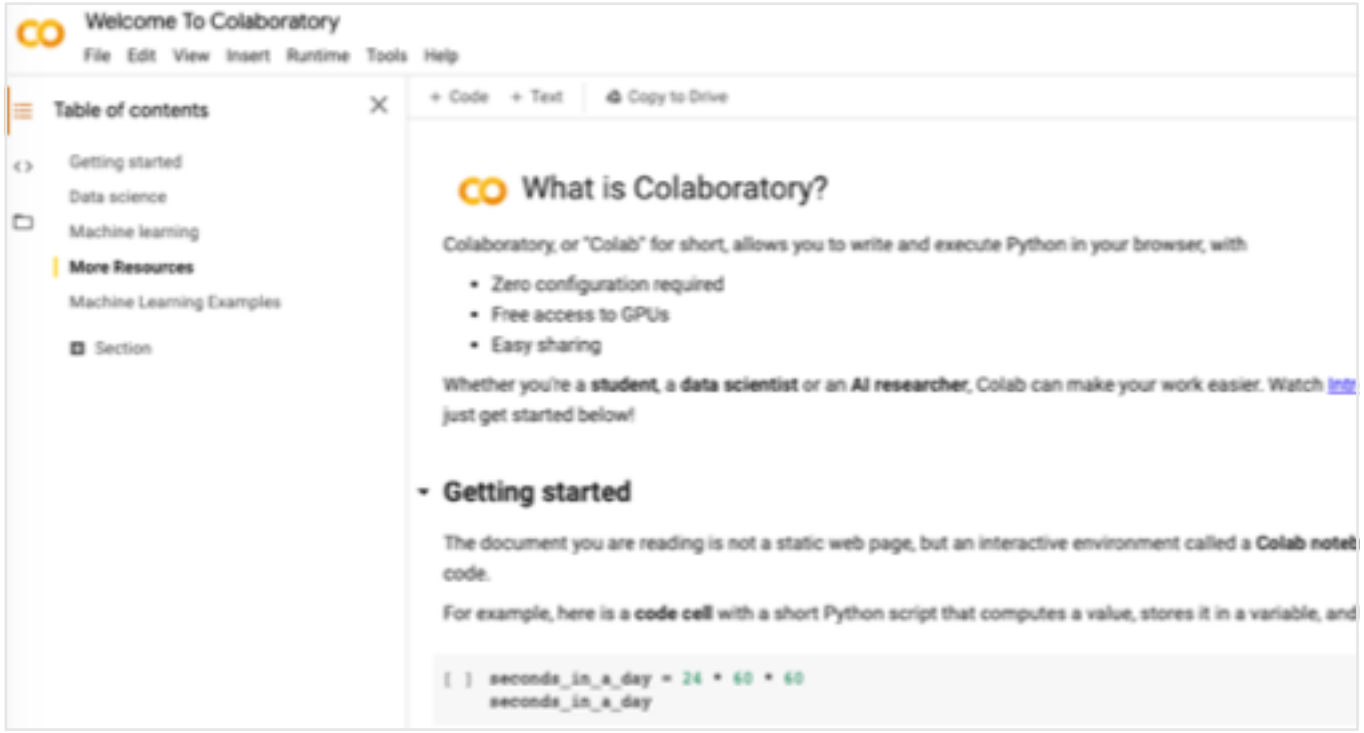
- Traditionally, we can install an IDE on our own computer in order to develop a program

Developing machine learning models

- Having an IDE on our own computer for developing computer programs is great, but it also has limitations:
  - You will need to download, install, and set up everything on your own computer
  - Difficult to develop and train computer programs that require a lot of computing resources (those programs that handle big geospatial data)

Google Colab

- Google Colab is a Jupyter Notebook style application that allows you to develop computer programs on the powerful Google Cloud Platform, for free!



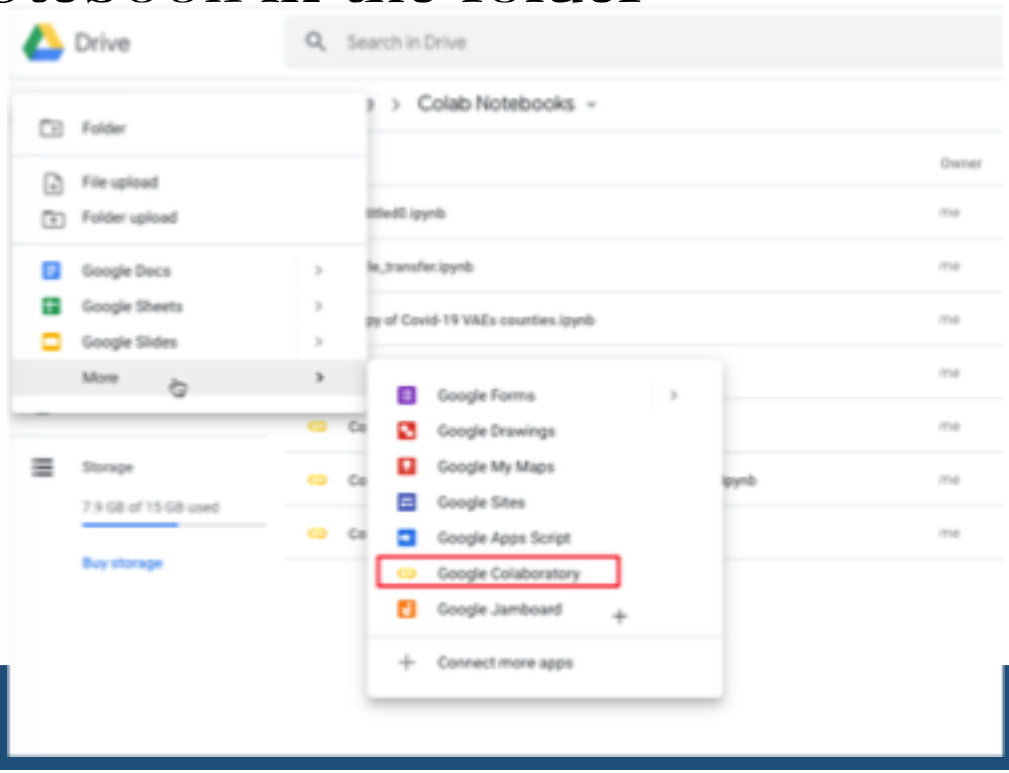
https://colab.research.google.com/notebooks/intro.ipynb

Google Colab

- To create a new notebook in Google Colab, you can directly create new notebook at the link below:

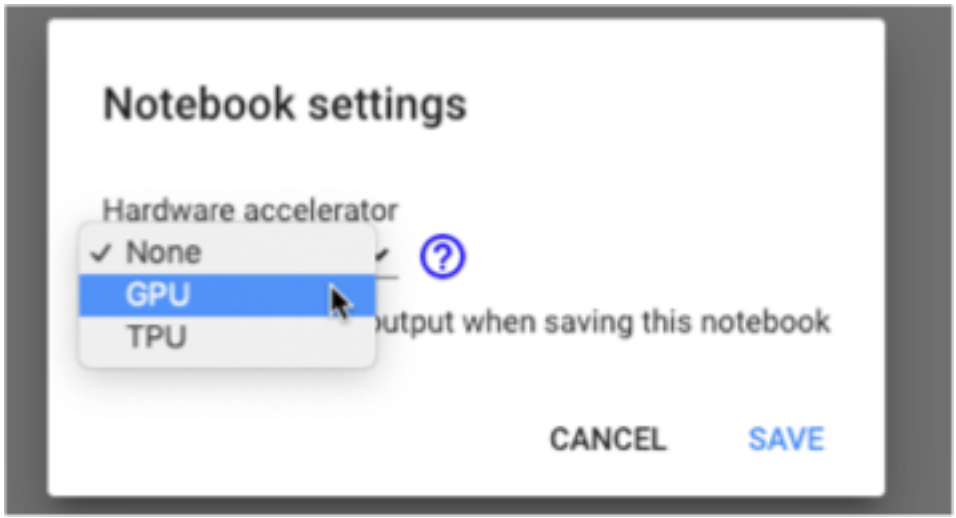
  https://colab.research.google.com/notebooks/intro.ipynb

- Or, you can first create a folder in your Google Drive, and then create a notebook in the folder



Google Colab

- Set up the runtime environment, and use Google's GPU or even TPU for free!
- You won't see much difference using or not using GPU, if you model is simple. However, it can make a huge difference for running DL models on big geospatial data

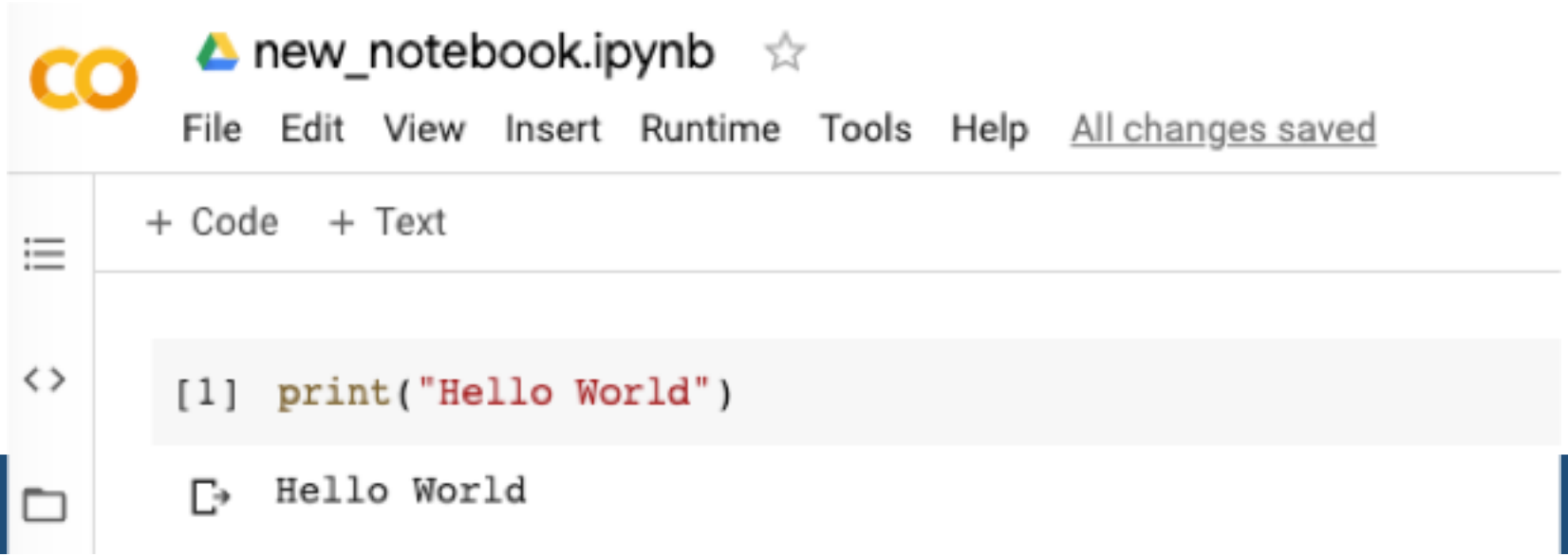


Google Colab

- Google Colab already have a basic working environment configured for you; How to know which version of Python is running on your Google Colab?

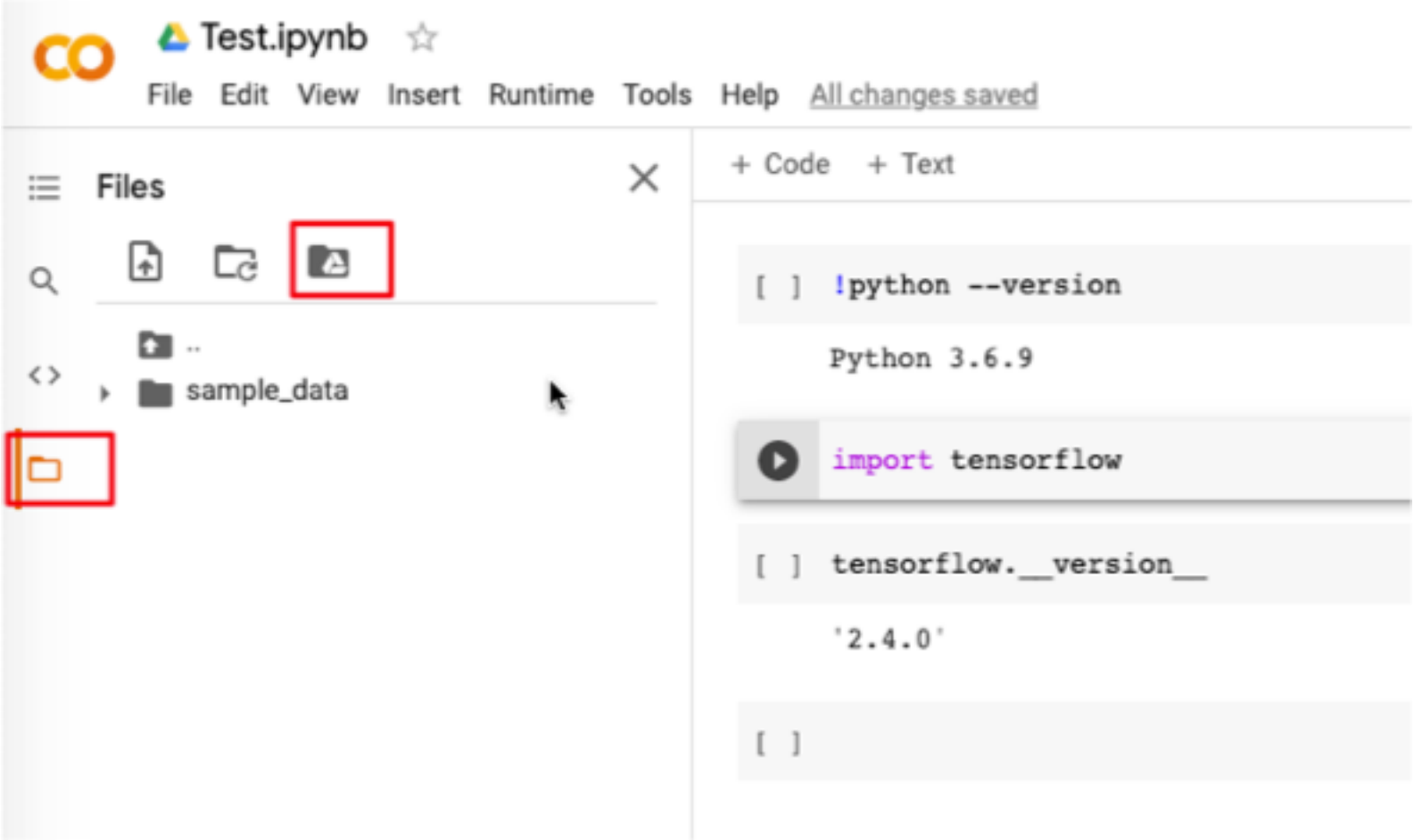
```
!python --version
```

- Running "Hello World" in Google Colab



Google Colab

- Connecting your Google Drive to Colab



Google Colab

- Google Colab is very cool! How exactly does it work?
  - When you connect your notebook to Google Colab, it sets up a virtual machine (like the virtual environment) on the cloud
  - This virtual machine comes with most Python libraries (and can be equipped with a GPU or TPU), which means you don't need to set up anything in order to run your code
  - You run your notebook in the virtual machine
  - After you have finished running your code, the virtual machine is DESTROYED!!! However, your code is saved in your Google Drive.

Google Colab

- What does this mean?
  - Google Colab is a free cloud-based resource for anyone, so it has to allocate its computer resources dynamically
  - A program can run maximum 12 hours in Google Colab but this time also varies depending on the concurrent usage of Google Colab
  - Your code is saved in your Google Drive, but you should save your work in progress from time to time, e.g., processed data, trained model… which will be destroyed once you disconnect.