Introduction to Google Colab

Vinay Bhaip

September 2019

1 Introduction

As machine learning algorithms require more and more computational power to be successful, it is important to find a way to overcome this barrier in order to create successful models. To combat this, Google has released Google Colab, an online coding environment with the ability to harness the power of GPUs, which are advanced processing units that vastly accelerate the time needed for machine learning models to learn.

2 Setting Up

After logging into your Google account, go to colab.research.google.com. Right off the bat, you'll have the ability to either import code from Google Drive, Github, or your own computer. You can also just create a new Python3 or Python2 notebook, though we recommend Python3 as Python2 will lose support in 2020.

It's important to note that when creating a file, it will be made with the .ipynb extension, which indicates its a Jupyter Notebook, rather than just a file. A Jupyter Notebook allows for some more nice features than a normal Python file. For the purposes of what we'll do, you can treat both as the same, however.

The most powerful feature of Google Colab is the ability to run your code with a lot of computational power. Under the "Runtime" tab, you'll see an option to "Change runtime type." When you click that, you'll have the ability to change your hardware accelerator to a GPU, which is ideal for running machine learning algorithms. Additionally, you'll see an option to use the very powerful TPU, a processing unit made by Google specifically tailored for AI and ML. It's harder to configure most ML libraries for TPUs than it is to configure them for GPUs, so we recommend just sticking with GPUs. Also keep in mind that Colab will time out eventually when you use the TPUs, so it is not ideal in all scenarios. In the beginning of the year, you probably will not require the use of a GPU for our competitions, but when we delve into neural networks and beyond, it will definitely be helpful.

3 Running Code

It's pretty simple to run code in Colab. It's structured just like a Jupyter Notebook, which means that you'd write code in "blocks" and execute those blocks of code. This is helpful with debugging and reading your code.

Most libraries you need such as Numpy, Keras, and Tensorflow are already installed into Google Colab so you don't need to worry about that. For reading in files, such as CSV files from our in-house Kaggle competitions, we recommend you check out https://colab.research.google.com/notebooks/io.ipynb, which covers this topic.

We encourage you to explore all the tools Google Colab has to offer as some may be helpful to whatever you are working on.