CS109b, Spring 2018 - Lab 7

David Wihl, CS109b S

3/8/2018

Intro

In this lab, we will

1. Set up Python on your local environment
2. Ensure that you can access JupyterHub for Python development
3. Try out Google Cloud

Part 1 - Python Setup

You likely have Python already on your machine. However, Python has multiple versions (e.g. 2.7 and 3.6), as well as multiple versions of libraries with multiple dependencies.

If you are already comfortable with configuring Python and associated libraries, feel free to follow your existing procedures. However, we will not be able to support non-standard configurations. You are on your own.

We will be using Python 3.6.

Here are the libraries we will be using:

```
numpy=1.14.1
scipy=1.0.0
matplotlib=2.1.2
Pillow=5.0.0
scikit-learn=0.19.1
keras=2.1.2
tensorflow=1.3.0
```

Later versions of these libraries should be fine.

We recommend using venv to manage your installation of Python and associated packages. The advantage of using venv is that you can keep your versions of Python and associated libraries consistent for a given project. If you've done several Python projects, you know that at some point, there will be mutually exclusive dependencies and something will break, badly.

Step 1:

Follow the tutorial at: https://docs.python.org/3/tutorial/venv.html

to create your first virtual environment.

Step 2:

Using the list of packages above, install the packages into your virtual environment.

Step 3:

Try one of the sci-kit learn examples, such as

http://scikit-learn.org/stable/auto_examples/plot_cv_predict.html#sphx-glr-auto-examples-plot-cv-predict-py

Step 4:

Ensure that Pandas (and implicitly numpy) is working.

https://pandas.pydata.org/pandas-docs/stable/10min.html

You don't have to do the entire tutorial. Going through "Object Creation" is sufficient.

Step 5:

Try Jupyter, with Keras.

Download the notebook and do part 2:

https://github.com/cs109/2018-ComputeFest/blob/master/ImageClassification.ipynb

(No need to do part 1 or 3)

Part 2 - JupyterHub

Every student has access to a JupyterHub instance that includes a GPU.

Launch JupyterHub from Canvas, upload the ImageClassification notebook and try running it on your JupyterHub instance. The GPU should be used.

Open a terminal window to your JupyterHub instance and run `nvidia-smi` to ensure that the GPU is installed, operational, and in use when you run the expensive training portion of the ImageClassification notebook. You can set a larger number of epochs if necessary to have it run for longer.

Part 3 - Google Cloud

This part is optional and may be useful if you have large datasets in the future.

Step 1: Sign up for Google Cloud

Go to https://console.cloud.google.com/start to create a new Google Cloud account. You should have a $300 credit.

You may need a credit card to enable billing for certain services. If you are prudent with your usage and do not leave instances running unnecessarily, you should not need to actually pay money. *This is your responsibility.*

Step 2: Try out Datalab

Datalab is Google's version of Jupyter running on their cloud.

Go to https://cloud.google.com/datalab/

A quickstart tutorial is available at https://cloud.google.com/datalab/docs/quickstart including how to enable GPUs.

Step 3: Try out Colaboratory

Colaboratory is a research version of Jupyter notebooks that work like Google Docs so there is real-time collaborative editing. GPU instances are also available for free.

Upload a Jupyter notebook to your Google Drive. Double-click to open it. You should have an option to open it with Colaboratory. If not, let us know. This project is still in research and may not yet work for everyone.

Step 4: Explore more of Google Cloud

You can waste days exploring. For example:

- Use TPUs (currently in beta): https://cloud.google.com/tpu/docs/
- Vision API (might be useful for some projects): https://cloud.google.com/vision/docs/quickstarts
- BigQuery (get thousands of machine to work on a single query) https://cloud.google.com/bigquery/docs/quickstarts