EE-559: Practical Session 14

this doc : https://goo.gl/KJTXLT

[version of: May 28, 2018]

Introduction

This practical session is about going through all the steps in a typical TensorFlow ML workflow : Getting the data, transforming it into TensorFlow Example datasets, and using the Estimator interface for training, evaluation, and exporting of the models.

The exercises of this practical session are contained in a single Colab: https://goo.gl/8uaMxk - Please click on that link now and then "**Save a Copy in Drive…**" (in the "File" menu).

This Colab contains a lot of code. In a first pass, you should try to walk through the "non optional" parts in a quick pace, only completing some small fragments where necessary. Later you should take some more time to play around with the different components to better understand what is going on, and to improve the performance of the (very basic) models. Most importantly, you probably want to come back to this practical session and use it as a starting point for your own ML problem.

1. Data

Examine the QuickDraw data, and convert it to a TensorFlow dataset of tf.train.Example protocol buffers, with sharded files for training, evaluation, and testing. Generating the dataset with the "img_64" feature will take ~2 minutes for 70k examples.

Optional : Generate the dataset for the RNN. To make the problem more interesting, you might to want to come back later and create larger datasets with more classes. Be aware that your VM only persists ~24h so you would want to either run this Colab on your computer (or on a VM), or move the dataset to the cloud (see "3. Cloud") once its generated.

2. Estimators

Use the data prepared in the previous section and train a couple of estimators. You will first use two canned estimators (LinearClassifier, DNNClassifier) to which you merely specify the feature columns, the configuration, and the input_fn for accessing the data. Then you will implement a custom estimator that implements a CNN.

Additional ideas (to solidify your TF skills after the lab)
- Find better parameters for the custom model.

- Implement and understand the RNN. Improve performance. You can get inspiration here:
  https://www.tensorflow.org/tutorials/recurrent_quickdraw
  https://github.com/tensorflow/magenta/tree/master/magenta/models/sketch_rnn
- Export the model as a saved model. Load it in Python and classify examples.
  https://www.tensorflow.org/api_docs/python/tf/estimator/Estimator#export_savedmodel
- Write a JS app that lets you draw sketches on the cellphone, and then loads the exported saved model to classify the sketches on-device.

3. Cloud

Beware : This section assumes basic sysadmin skills (connect to computers via SSH, register additional SSH keys, install software packages, copy files over SCP, ...) that won't be covered in detail.

This section contains little code but a lot of practical instructions. After moving the code out of Colab and into its own binary that you can run the training on your own computer. Then you'll have to create a cloud account, enable different APIs and then run the model on a VM in the cloud and via the Cloud ML engine.