

**From Keras to Plotly: Building Deep Learning Models and Visualizing Predictions**

**UVA DataPalooza, November 9, 2018 Machine Learning Skills Session**

**1:15 – 2:45 PM**

*By members of the Booz Allen Women in Data Science Group and the Cakes & Tensors Booz Allen AI Meetup*

*Catherine Ordun and Michael Rushin*

For any technical difficulties you may face during the skills session, Booz Allen data scientists Sian, Jessica, Mike, and Michael will be on stand-by to help you out.

**Keras Session: 30 - 40 minutes (Catherine)**

In this tutorial you’ll be learning how to build a multi-label text classification model using data from Stack Overflow. A multi-label problem differs from a multi-class problem, in that any sample can be assigned to multiple classes, as opposed to a standard binary classification problem where one sample is assigned to only specific class. In this tutorial you’ll learn what about the Word2Vec Skip Gram implementation and how to develop a word embedding layer for a Long Short-Term Memory (LSTM) recurrent neural network. The predictions we make will be used to visualize in Plotly.

We’ll embark on our Keras journey together using a Jupyter Notebook file loaded in the new Google Colaboratory environment that allows for 12 hours of free GPU run-time on a Tesla K80. Since we don’t have a lot of time together, I’ve pre-run several of the models we’ll use ahead of time, so you can just load them up and run them during the tutorial. I’ve also written a processing script that, for future use, you can modify for your customization.

For this session you don’t need anything more than your laptop and a Google account in order to access your Google Drive. Everything you need is located on this publicly accessible Google Drive Folder. Since this is connected to my own personal drive, please do not share this link outside of the tutorial. It will be shut down after the tutorial is finished.

Step 1- Log into your personal Google Drive account (like how you access your gmail). *You must be using Chrome, or you may encounter Google Colaboratory errors.*

Step 2 – Click on the link below to access the ‘Tutorial’ folder.

|  |
| --- |
| <https://drive.google.com/drive/folders/1Ewmbi_HEQ62UcAaNahLGT0WQJB2yoocG?usp=sharing> |

Step 3 – Right click on the folder and “Add to your local drive”.

Step 4 – Now, right click on the “.ipynb” file. This is the Jupyter Notebook. When you right click, do you see “Colaboratory”?

Step 5a – If yes, go ahead and click it and the notebook will open inside Colaboratory for you.

Step 5b – If not, continue to right click and “Connect more apps”

Step 6 – The Google apps frame will open, search for “Colaboratory” in the upper right hand corner search bar. Remember, it’s one ‘l’, not ‘ll’ (two).

Step 7 – Follow the instructions and “connect” to it, next to the “share” icon. Give it a second to update.

Step 8 – Go back to Step 5a – right click on the “.ipynb” file and open in Colaboratory.

*See below for additional screenshots:*



**Note:** Feel free to look through the notebook. **But DO NOT RUN THE ENTIRE NOTEBOOK. Please run the first few cells to test mounting your drive all the way down to importing the csv, but please do not run more than that.** The functions and orchestration of the notebook, which I will guide you through during the tutorial, has been carefully calibrated in order to not run out of memory. Although the Google Colaboratory is free, there is restricted memory and running multiple operations at once will result in a runtime error (it’s not totally free!). Therefore, I recommend you wait until the skills session so we can walk through the notebook together.

**Plotly Session: 30 - 40 minutes (Michael)**

In this session you will be learning how to create a web application for data visualization using Plotly Dash. We will use the results of the predictions we made from the neural net as a csv file and load them into a Jupyter Notebook to build the dashboard. The code is located on my public Github account provided below. You need to install all the local dependencies and set up a virtual environment in Steps 1 and 2 below. Then, install in Step 3 the Github repository and then open the Jupyter Notebook. I will show you how to build a table and a scatterplot.

1. INSTALL ANACONDA
   1. The open source Anaconda platform is the industry standard for performing data science tasks in Python. Please follow the instructions for your operating system from the Anaconda website, <https://www.anaconda.com/>.
2. CREATE VIRTUAL ENVIRONMENT
   1. Once Anaconda is installed, create a virtual environment for the application. A virtual environment allows us to isolate the Python packages installed for Anaconda and is good practice.
   2. Open Anaconda Navigator
   3. Select the “Environments” tab
   4. Select the “Create” button
   5. Do not alter the default settings
   6. Title the environment “uva-datapalooza”
   7. Select the “Create” button
3. INSTALL APPLICATION REQUIREMENTS
   1. Navigate to <https://github.com/mrushin/UVA-DataPalooza>
   2. Download the zip file
   3. Open the Anaconda Prompt and change directories to the UVA-DataPalooza folder
   4. Type “conda activate uva-datapalooza”
   5. Type “pip install -r requirements.txt”
4. INSTALL JUPYTER NOTEBOOK KERNEL
   1. Jupyter Notebook is an interactive computational environment, in which one can combine code execution, rich text, mathematics, plots and rich media. The data processing, analytics, and visualization notebooks (.ipynb files) are can be run using the following steps.
   2. Open the Anaconda Prompt and change directories to the application folder
   3. Type “conda activate uva-datapalooza”
   4. Type ‘python -m ipykernel install --user --name uva-datapalooza --display-name "Python (uva-datapalooza)"’ NOTE: Only required once.
   5. Type “jupyter notebook”
   6. Navigate to desired notebook and open
   7. From the top menu, select kernel, Restart & Run All