

Homework 3: Prerequisites

COMS W4705_001: Natural Language Processing

Prof. Kathleen McKeown, Fall 2017

To begin, make sure you have an instance with GPU with CUDA setup, refer to ["https://cloud.google.com/compute/docs/gpus/add-gpus"](https://cloud.google.com/compute/docs/gpus/add-gpus) . You would need one <NVIDIA Tesla K80> GPU, at least 16 GB (Preferably 20 GB) of RAM, and at least 75 GB of disk space. To reiterate, the homework needs to be programmed in Python3 and using Keras and Tensorflow-Gpu libraries.

Most importantly, make sure you **STOP YOUR INSTANCE** when you are not using it!

Next, use "sudo pip3 install jupyter (or pip if python3 is default)" command to install Jupyter, as for this assignment you will use Jupyter Notebook, set up with google cloud instance. To set up jupyter notebooks on GCP, refer this resource ["https://medium.com/towards-data-science/running-jupyter-notebook-in-google-cloud-platform-in-15-min-61e16da34d52"](https://medium.com/towards-data-science/running-jupyter-notebook-in-google-cloud-platform-in-15-min-61e16da34d52). You could also use ["http://cs231n.github.io/gce-tutorial/"](http://cs231n.github.io/gce-tutorial/) (it's a slightly old version), from section titled "Using Jupyter Notebook with Google Compute Engine". (Do remember in Targets select "All instances in the network" for Firewall rules.)

Once you run "jupyter-notebook --no-browser --port=<PORT-NUMBER> ", your terminal will show something like

"Copy/paste this URL into your browser when you connect for the first time, to login with a token:

<http://localhost:7000/?token=262f17d25c9d323c21c2dfca44e3609969ae8ebcf5432cf7>"

Replace "localhost" with your static GCP instance's IP and paste on the browser to access the Jupyter Notebook. If you are unfamiliar with Jupyter notebooks, ["https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/execute.html"](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/execute.html) is a good resource, but you will find the GUI very intuitive and straightforward. Another good resource is ["http://cs231n.github.io/ipython-tutorial/"](http://cs231n.github.io/ipython-tutorial/). You can execute cells serially and if execution fails at

one cell, it still keeps in memory the previously executed cells, which will be very helpful for the Homework.

Once you are finished with that, run
<wget <http://nlp.stanford.edu/data/glove.42B.300d.zip>> to get pretrained word embeddings, and extract them using the “unzip” command (you might have to install unzip).

Please see the pinned post on Piazza about “(Homework 3) Environment Setup” if you have any doubts or questions regarding the set-up.

Again, make sure you **STOP YOUR INSTANCE** when you are not using it!