# COMS W4705 - Natural Language Processing - Homework 5

**Due: 11:59pm on Friday, December 28th. NO LATE DEADLINE OR EXTENSIONS.Total Points: 100**

Working on homework 5 is **optional.** The homework component of your final grade for this course will be determined based on the 4 highest homework scores. This means that, if you work on homework 5, your lowest homework score will be dropped.

While you should be able to train the image caption generator on your own machine (i.e. on a system without a GPU), we strongly recommend that you try to use a GPU instance on the Google Cloud. See the instructions for obtaining free credits and setting up such an instance below.

# Programming Component - Image Captioning (100 pts)

Please make sure you are developing and running your code using Python 3.

**Introduction**

In this assignment you will construct a basic image captioning system using an LSTM encoder-decoder model. We will also use a pre-trained Convolution Neural Network (CNN) trained on an object detection task to compute image embeddings.

## Prequisites: Installing necessary packages

You should already have Keras, Tensorflow, and numpy installed from homework 3. You will also need matplotlib, as well as PIL (Python Image Library, specifically the pillow (Links to an external site.)Links to an external site. fork.)

For this assignment, you will use Jupyter Notebook (Links to an external site.)Links to an external site..

We suggest that you use the Python package management system pip. On most systems, the following commands will work:

```
$ pip install matplotlib
$ pip install pillow
$ pip install jupyter
```

## Running the Jupyter Notebook

Download the notebook homework5.ipynb

Then open it in Jupyter Notebook like this:

```
$ jupyter-notebook homework5.ipynb
```

All further instructions are included in the notebook.

## Free stuff: Complementary Google Cloud Access

Courtesy of a Google Cloud Platform (GCP) Education Grant, you can get $50 of free GCP credits! You can use these credits to create VM instances, including GPU instances.

If you do not have a GCP account, you can set up an account here, which also provides some initial credits. https://cloud.google.com/free/ (Links to an external site.)Links to an external site.

You may have to use your Columbia email so you can link the $50 education grant credits to your account.

Here is the URL you will need to access in order to request a Google Cloud Platform coupon. You will be asked to provide your school email address and name. An email will be sent to you to confirm these details before a coupon is sent to you.

Student Coupon Retrieval Link (Links to an external site.)Links to an external site.

(Links to an external site.)Links to an external site.

- The page says COMS W4701. Don't worry about this (I applied for the education grant when I was still planning to teach 4701 this Fall).
- You will be asked for a name and email address, which needs to match the domain (columbia.edu). A confirmation email will be sent to you with a coupon code.
- You can request a coupon from the URL and redeem it until: 1/4/2019
- Coupon valid through: 9/4/2019
- You can only request ONE code per unique email address.

## Getting Started with GCP

**BIG REMINDER: Make sure you stop your instances! (taken from Stanford's cs231n Course)****!This is super important!**

Don’t forget to **stop your instance** when you are done (by clicking on the stop button at the top of the page showing your instances), otherwise you will **run out of credits** and that will be very sad. :(

If you follow our instructions below correctly, you should be able to restart your instance and all software will still be available.

### Requesting a quota increase

By default the quota for GPU instances you can spin up is 0. You can request a quota increase as follows:

- In the main menu in the GCP console, go to "IAM & admin" and then select "quotas".
    - Under "Quota type" select "all quotas". Under "Metric", first click on "None" and then select "NVidia P100 GPUs". Under "location" select the availability zone that you would like to create you instance in. For example, my default is us-east1.
    - Select the service that shows up. Then click on "+ edit quotas" (near the top).
        - Fill in the form on the right to request a new limit of 1. Enter a description, for example "train neural network models for a course I am taking at Columbia University".
        - Then click "Submit Request".
- You should receive a confirmation email telling you that the quota increase may take up to two business days. Typically this should happen way faster.

### Creating a GPU instance

- Once you have an account, go to Compute Engine.

- Select VM instances in the menu on the left.
- Click on "create new instance".
    - Under "machine type" select 8 vCPUs, 30GB memory.
    - Click on "customize"
        - Under "GPU" select 1.
    - Under "boot disk" click on "change".
        - Scroll down and select "Intel® optimized Deep Learning Image: TensorFlow 1.12.0 m14 (with Intel® MKL-DNN/MKL and CUDA 10.0)"
    - Scroll down and click on "create"
- **Important: Make sure to stop the instance once you are done working with it. See warning above.**

### Logging into the instance

- Click on "ssh" next to the instance, or use your favorite ssh client (more info about connecting to the instance (Links to an external site.)Links to an external site.)
- Follow the prompts for installation of the nvidia drivers.
- You now have a running instance. Check that the GPU is working by running "nvidia-smi".
- All necessary packages are already in the image.

### Downloading the homework notebook and running jupyter

- First click on the GCP main menu (horizontal bars on the top left) and choose "VPC network", then "firewall rules".
    - Click on "create firewall rule"
    - choose any name like "allow port 5000"
    - under "targets" select "all instances in network"
    - in "source ip ranges" enter "0.0.0.0/0"
    - under "protocols and ports" check tcp and enter 5000.
- On your instance, download the notebook using "wget http://www.cs.columbia.edu/~bauer/4705/homework5.ipynb"
- Run "jupyter notebook --generate-config"
- Edit ~/.jupyter/jupyter_notebook_config.py using your favorite text editor. Add the line

    ```
    c.NotebookApp.ip = '0.0.0.0'
    c.NotebookApp.allow_remote_access = True
    ```

    This will make the notebook server listen to incoming web requests. Close the file.
- Run the notebook server "jupyter notebook --no-browser --port=5000"
- Copy the URL that shows up, and paste it into a browser address bar. Change the domain name to your instance's public IP.
- You should now be able to open the notebook.

**REMINDER: Make sure you stop your instances!**