This assignment is due on **Wednesday, April 22 2020** at 11:59pm PST.

Goals

In this assignment you will practice putting together a simple image classification pipeline based on the k-Nearest Neighbor or the SVM/Softmax classifier. The goals of this assignment are as follows:

- Understand the basic **Image Classification pipeline** and the data-driven approach (train/predict stages)
- Understand the train/val/test **splits** and the use of validation data for **hyperparameter tuning**.
- Develop proficiency in writing efficient **vectorized** code with numpy
- Implement and apply a k-Nearest Neighbor (**kNN**) classifier
- Implement and apply a Multiclass Support Vector Machine (**SVM**) classifier
- Implement and apply a **Softmax** classifier
- Implement and apply a **Two layer neural network** classifier
- Understand the differences and tradeoffs between these classifiers
- Get a basic understanding of performance improvements from using **higher-level representations** as opposed to raw pixels, e.g. color histograms, Histogram of Gradient (HOG) features, etc.

Setup

You can work on the assignment in one of two ways: **remotely** on Google Colaboratory or **locally** on your own machine.

Regardless of the method chosen, ensure you have followed the setup instructions before proceeding.

Option A: Google Colaboratory (Recommended)

Download. Starter code containing Colab notebooks can be downloaded here.

If you choose to work with Google Colab, please watch the workflow tutorial above or read the instructions below.

1. Unzip the starter code zip file. You should see an `assignment1` folder.
2. Create a folder in your personal Google Drive and upload `assignment1/` folder to the Drive folder. We recommend that you call the Google Drive folder `cs231n/assignments/` so that the final uploaded folder has the path `cs231n/assignments/assignment1/`.
3. Each Colab notebook (i.e. files ending in `.ipynb`) corresponds to an assignment question. In Google Drive, double click on the notebook and select the option to open with `Colab`.
4. You will be connected to a Colab VM. You can mount your Google Drive and access your uploaded files by executing the first cell in the notebook. It will prompt you for an authorization code which you can obtain from a popup window. The code cell will also automatically download the CIFAR-10 dataset for you.
5. Once you have completed the assignment question (i.e. reached the end of the notebook), you can save your edited files back to your Drive and move on to the next question. For your convenience, we also provide you with a code cell (the very last one) that automatically saves the modified files for that question back to your Drive.
6. Repeat steps 3-5 for each remaining notebook.

Note 1. Please make sure that you work on the Colab notebooks in the order of the questions (see below). Specifically, you should work on kNN first, then SVM, the Softmax, then Two-layer Net and finally on Image Features. The reason is that the code cells that get executed *at the end* of the notebooks save the modified files back to your drive and some notebooks may require code from previous notebook.

Note 2. Related to above, ensure you are periodically saving your notebook (`File -> Save`), and any edited `.py` files relevant to that notebook (i.e. **by executing the last code cell**) so that you don't lose your progress if you step away from the assignment and the Colab VM disconnects.

Once you have completed all Colab notebooks **except** **`collect_submission.ipynb`**, proceed to the submission instructions.

Option B: Local Development

Download. Starter code containing jupyter notebooks can be downloaded here.

Install Packages. Once you have the starter code, activate your environment (the one you installed in the Software Setup page) and run `pip install -r requirements.txt`.

Download CIFAR-10. Next, you will need to download the CIFAR-10 dataset. Run the following from the `assignment1` directory:

```
cd cs231n/datasets
./get_datasets.sh
```

Start Jupyter Server. After you have the CIFAR-10 data, you should start the Jupyter server from the `assignment1` directory by executing `jupyter notebook` in your terminal.

Complete each notebook, then once you are done, go to the submission instructions.

Q1: k-Nearest Neighbor classifier (20 points)

The notebook **knn.ipynb** will walk you through implementing the kNN classifier.

Q2: Training a Support Vector Machine (25 points)

The notebook **svm.ipynb** will walk you through implementing the SVM classifier.

Q3: Implement a Softmax classifier (20 points)

The notebook **softmax.ipynb** will walk you through implementing the Softmax classifier.

Q4: Two-Layer Neural Network (25 points)

The notebook **two_layer_net.ipynb** will walk you through the implementation of a two-layer neural network classifier.

Q5: Higher Level Representations: Image Features (10 points)

The notebook **features.ipynb** will examine the improvements gained by using higher-level representations as opposed to using raw pixel values.

Submitting your work

Important. Please make sure that the submitted notebooks have been run and the cell outputs are visible.

Once you have completed all notebooks and filled out the necessary code, there are **two** steps you must follow to submit your assignment:

1. If you selected Option A and worked on the assignment in Colab, open `collect_submission.ipynb` in Colab and execute the notebook cells. If you selected Option B and worked on the assignment locally, run the bash script in `assignment1` by executing `bash collectSubmission.sh`.

This notebook/script will:

- Generate a zip file of your code (`.py` and `.ipynb`) called `a1.zip`.
- Convert all notebooks into a single PDF file.

Note for Option B users. You must have (a) `nbconvert` installed with Pandoc and Tex support and (b) `PyPDF2` installed to successfully convert your notebooks to a PDF file. Please follow these installation instructions to install (a) and run `pip install PyPDF2` to install (b). If you are, for some inexplicable reason, unable to successfully install the above dependencies, you can manually convert each jupyter notebook to HTML (`File -> Download as -> HTML (.html)`), save the HTML page as a PDF, then concatenate all the PDFs into a single PDF submission using your favorite PDF viewer.

If your submission for this step was successful, you should see the following display message:

```
### Done! Please submit a1.zip and the pdfs to Gradescope. ###
```

2. Submit the PDF and the zip file to Gradescope.

Note for Option A users. Remember to download `a1.zip` and `assignment.pdf` locally before submitting to Gradescope.