



26.09.2022

Department of Computer Engineering
Prof. Dr. Hazım Kemal EKENEL

BLG 506E COMPUTER VISION ASSIGNMENT 1 k-Nearest Neighbor (kNN)

0. For this assignment and the others you will be given Stanford University CS231n course (<http://cs231n.stanford.edu/>) assignments. As stated in CS231n, you should be good at *Python*. Please have a look at *Python/NumPy/IPython* tutorials at <http://cs231n.github.io/>. Also we recommend you to have a *Linux* OS either locally (on your machine) or virtually (on your machine or Cloud services). Similarly, it is recommended to build a *Python* environment preferably by one of the methods below.

* *Anaconda* (<https://www.anaconda.com/>)

* *Miniconda* (<https://docs.conda.io/en/latest/miniconda.html>),

* *virtualenv* (<https://virtualenv.pypa.io/en/latest/>)

Check setup instructions page of CS231n (<http://cs231n.github.io/setup-instructions/>)

All works must be your own!

In your submission (.zip), provide all source files (.py, .ipynb etc.) that you used and your report (.pdf).

You should have comments in your code.

In your report, you should explain all the steps in detail.

You can also add comments to code or insert text cells for explaining the steps instead of preparing a report in .pdf format.

1. Download assignment 1 from Stanford's CS231n:

<http://cs231n.github.io/assignments2019/assignment1/>

You will be following knn.ipynb notebook.

2. Download and extract CIFAR-10 dataset using get_datasets.sh script under cs231n/datasets .

3. Start notebook with jupyter-notebook command. In the notebook, run library/parameter setups and load CIFAR10 dataset . You should have output like

Training data shape: (50000, 32, 32, 3)

Training labels shape: (50000,)

Test data shape: (10000, 32, 32, 3)

Test labels shape: (10000,)

(If you encounter Cannot import scipy.misc.imread error, check if your environment have scipy 1.3.0. If so, downgrade it to scipy<=1.2.0 or you can change the line "from scipy.misc import imread" to "from imageio import imread" to read images from different library in the file "cs231n.data_utils" -the second option is recommended for Colab user since it is hard to change some libraries in Colab environment)

4. Run visualizing part and see your images and classes.

5. Subsample your data with 5000 training samples and 500 test samples.
6. Initialize KNearestNeighbor class and train your dataset. Remember that training is just saving the dataset as variables.

(If you encounter No module named 'past' error, install future library to your environment)

7. Implement compute_distances_two_loops function in cs231n/classifiers/k_nearest_neighbor.py that computes Euclidean distance (L2 distance/norm) between test and training samples. Call this function from notebook.
8. Plot the Euclidean distance matrix between the training and the test set. Describe the plot. (Answer inline question 1)
9. Implement predict_labels function and calculate the accuracy of your classifier.
10. Answer inline question 2. (in the 5th step, answer theoretically)
11. Implement compute_distances_one_loops and compute_distances_no_loops functions. Call these from notebook. Compare all three implementations by their execution time.
12. Perform cross-validation with the training set split in 5 folds, and assert the best $k=[1, 3, 5, 8, 10, 12, 15, 20, 50, 100]$. Plot the cross-validation accuracy vs. k . Evaluate the number of correct predictions for the test set by using the best k .
13. Answer inline question 3.

For any question or discussion, you can e-mail to saritas21@itu.edu.tr or you can (before informing is preferable) and come to the SiMiT Lab (office no: 4105).

For Colab users, you can insert the code below (do not forget changing the "FOLDERNAME") to the beginning of the knn.ipynp file to mount Google Drive;

```
# This mounts your Google Drive to the Colab VM.
from google.colab import drive
drive.mount('/content/drive', force_remount=True)

# Enter the foldername in your Drive where you have saved the unzipped
# assignment folder, e.g. 'cs231n/assignments/assignment1/'
FOLDERNAME = None
assert FOLDERNAME is not None, "[!] Enter the foldername."

# Now that we've mounted your Drive, this ensures that
# the Python interpreter of the Colab VM can load
# python files from within it.
import sys
sys.path.append('/content/drive/My Drive/{}'.format(FOLDERNAME))

# This downloads the CIFAR-10 dataset to your Drive
# if it doesn't already exist.
%cd drive/My\ Drive/$FOLDERNAME/cs231n/datasets/
!bash get_datasets.sh
%cd /content/drive/My\ Drive/$FOLDERNAME
```