1. [Introduction]

In this assignment we conducted clustering using KNN and K- means on the SVHN dataset, which consists of digit images.

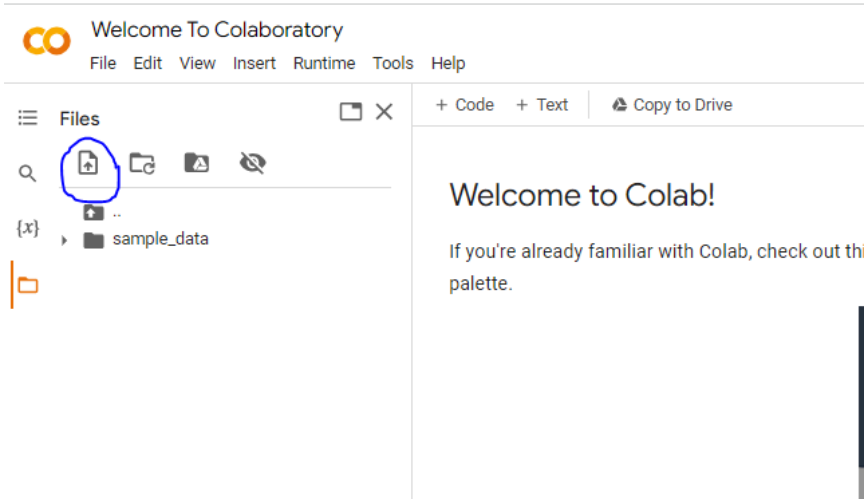
2. [Installation]

All the required libraries are imported in the ipynb file.

Note: If another IDE is used then please install the libraries.
pip install scikit-learn pandas numpy

3. [Usage]

- Open https://colab.google/
- Click on the open collab button.
- Click on upload and then click on browse.
- Select the 'train_32x32.mat' and ‘test_32x32.mat’ file and upload it.
- Click on 'upload to session storage'. Please do not upload it in any other folder.



- Run all the cells in the order.

4. [Data]

- The SVHN dataset, which consists of digit images. It has 10 classes, 1 for each digit. Digit '0' has label ‘0’, '1' has label 1, '9' has label 9. 73257 digits for training, 26032 digits for testing.

- Here, we have considered a subset of the training data with randomly chosen 2500 samples from each class from the original training set, totaling to 25,000 samples as my new training data. Test data remains the same with 26032 samples.