**CS498 AMO Homework 4**

Team :

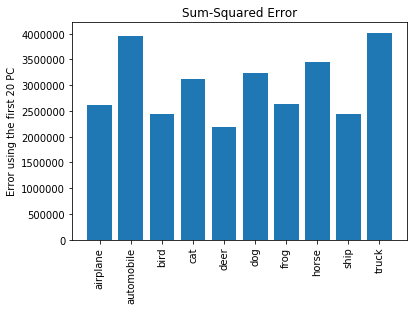
Minyuan Gu ([minyuan3@illinois.edu](mailto:minyuan3@illinois.edu), netid minyuan3)

Yanislav Shterev ([shterev2@illinois.edu](mailto:shterev2@illinois.edu), netid shterev2)

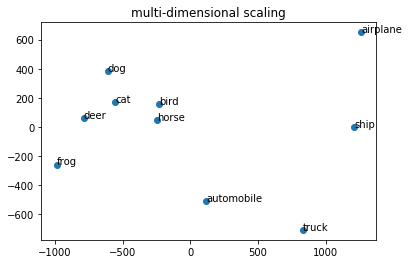
1. Page 1: **(10 points)** A plot of the mean image for each class.



1. Page 2: **(15 points)** A plot of the average sum of squared pixel-wise difference between the reconstructed and original images of each class **as a bar graph** vs. the class label.

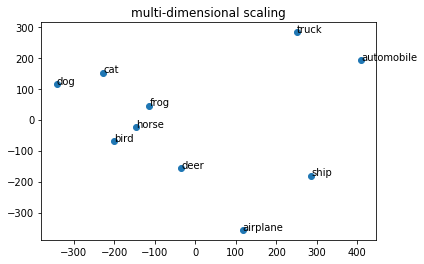


1. Page 3: **(25 points)** The 2D scatter plot obtained after performing principal coordinate analysis using Euclidean distance.



1. Page 4: **(25 points)** The 2D scatter plot obtained after performing principal coordinate analysis using the similarity metric in part C.

It is noticeable that there is a difference between the plot of the differences of mean images per category and the differences between the mean and the first 20 PC per category. The main difference is that the latter plot does not project the distances as accurate as the first one due to the fact it only uses the first 20 principal components. It is easy to see how classes like bird, horse and frog are considered close to one another because of this approximation omitting the rest features. Due to the scaling to lower space the absolute values of the distances is also smaller than these in the first plot- using only the means.



**Libraries used & Reference:**

**David Forsyth’s book** - Applied Machine Learning

**Trevor Walker’s lecture and sample code** – CS-498 Lecture videos

**csv** – for reading data from csv format: <https://docs.python.org/3/library/csv.html>

**CIFAR-10 dataset** - <https://www.cs.toronto.edu/~kriz/cifar.html>

**Numpy** - <http://www.numpy.org/>

**matplotlib** - to plot the sum-squared error, mean images and distances plots: <https://matplotlib.org/>

**scipy.spatial.distance. squareform and pdist**– to calculate Euclidean distance <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.spatial.distance.squareform.html>

**pickle** – to load the CIFAR-10 dataset: <https://docs.python.org/3/library/pickle.html>

**sklearn.decomposition**- PCA <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>