

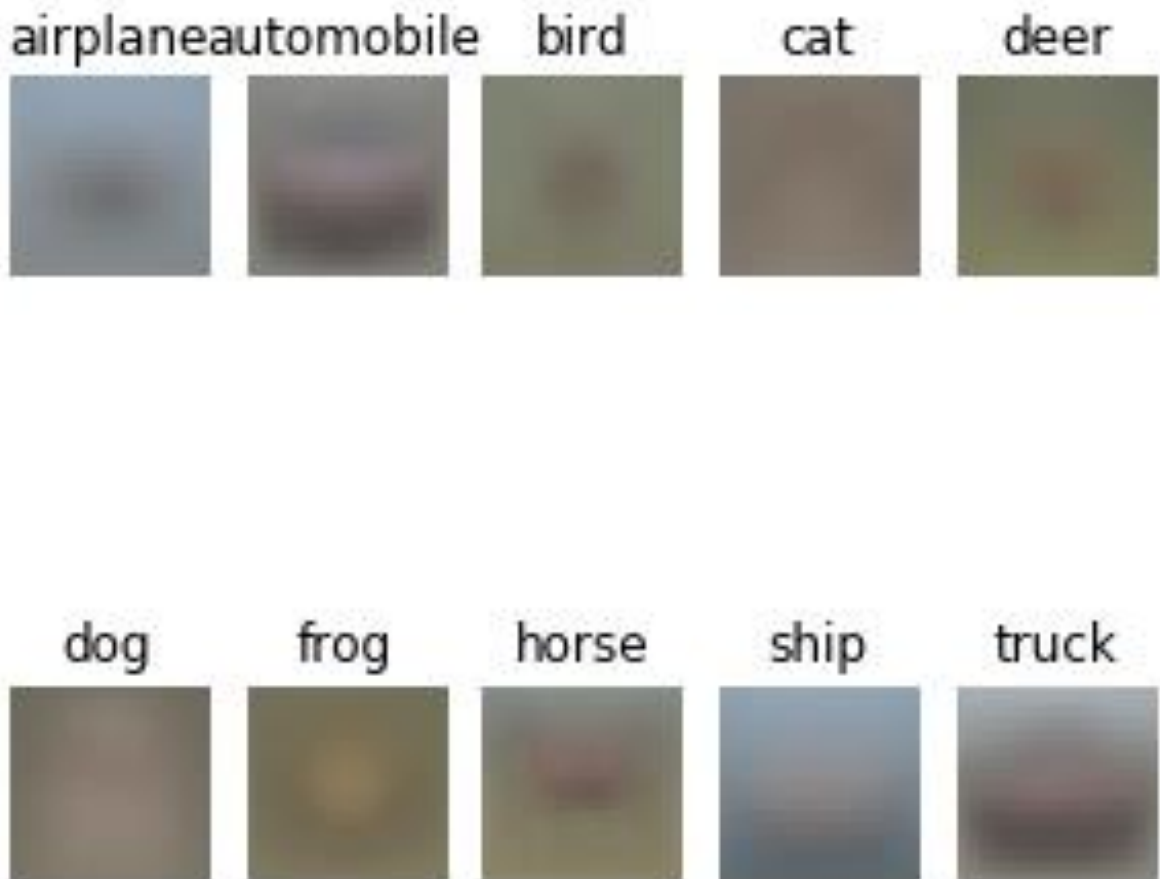
CS498 AMO Homework 4

Team :

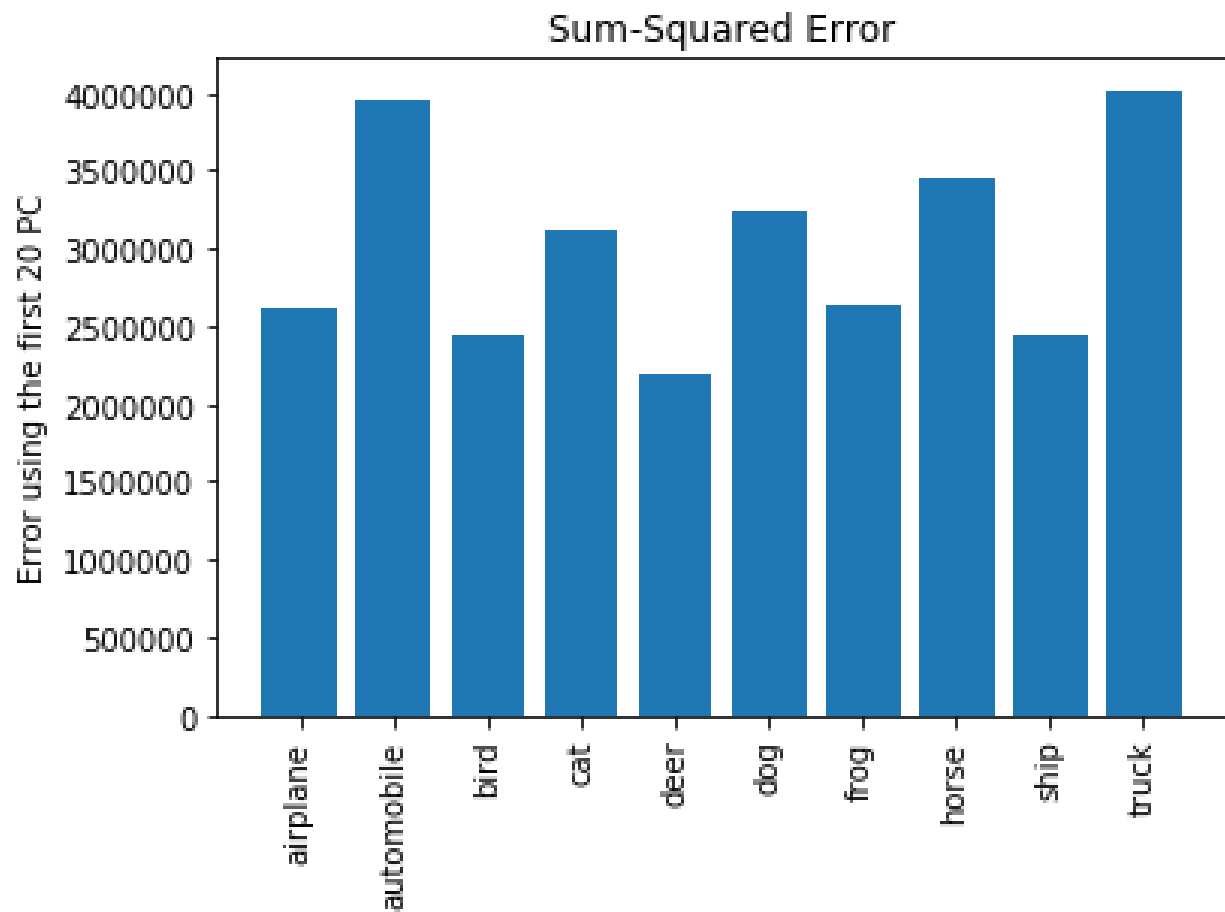
Minyuan Gu (minyuan3@illinois.edu, netid minyuan3)

Yanislav Shterev (shterev2@illinois.edu, netid shterev2)

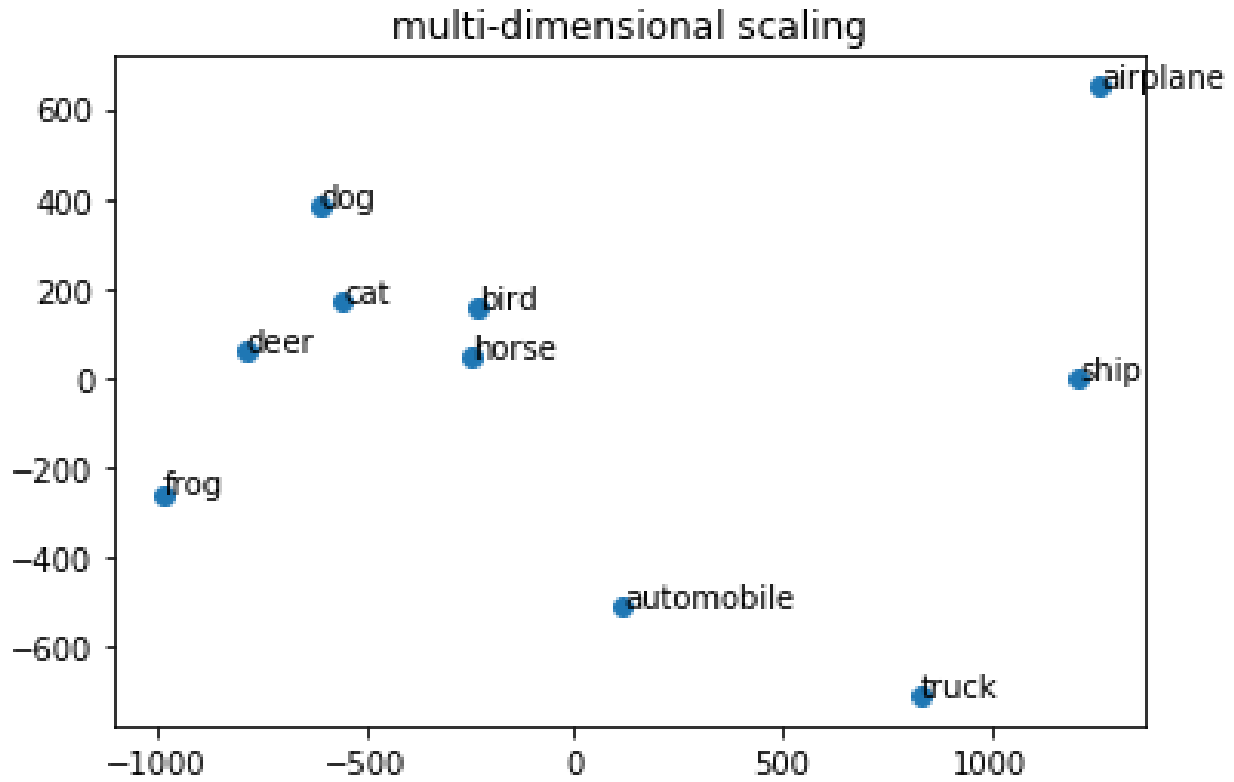
1. Page 1: **(10 points)** A plot of the mean image for each class.



2. 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.



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



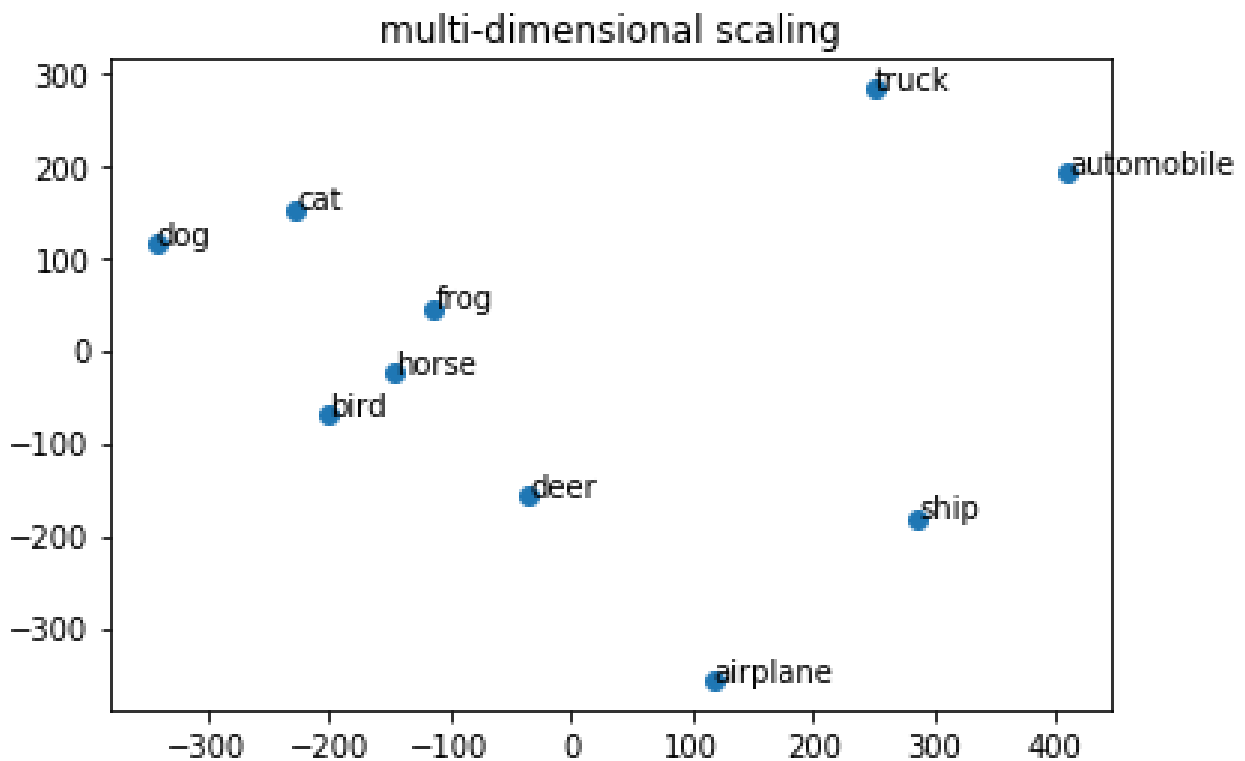
4. Page 4: **(25 points)** The 2D scatter plot obtained after performing principal coordinate analysis using the similarity metric in part C.

There are differences between the two plots. The reason for these differences is because the definitions of 'distance' are different.

- 1st Plot uses Euclidean distances between mean images per category
- 2nd Plot uses the $(E(A|B)+E(B|A))/2$ by swapping the first 20 PC from other category with its own mean and then calculate the difference between reconstructed images and original images per category.

Worth to mention, the 2nd plot distances between the same categories ($E(A|A) \neq 0$, diagonal of the distance matrix) are not zero due to the PCA projection from a different category, while for the 1st plot diagonal of D is zero.

Even the positions of the points are different between 2 plots, however, we see similar clustering in both plots: animal pictures tend to stay closer while other vehicles tend to stay far apart.



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-](https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.spatial.distance.squareform.html)

[0.14.0/reference/generated/scipy.spatial.distance.squareform.html](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-](https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html)

[learn.org/stable/modules/generated/sklearn.decomposition.PCA.html](https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html)