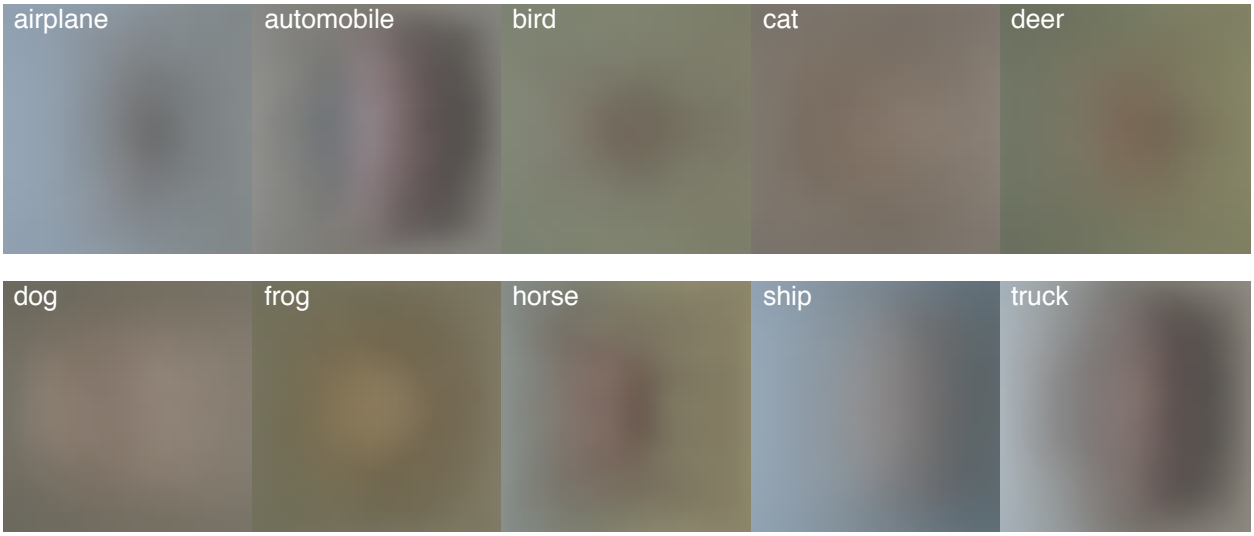
CS 498 AML: Homework 3

Rauhul Varma (rvarma2), Tadas Aleksonis (alekson2)

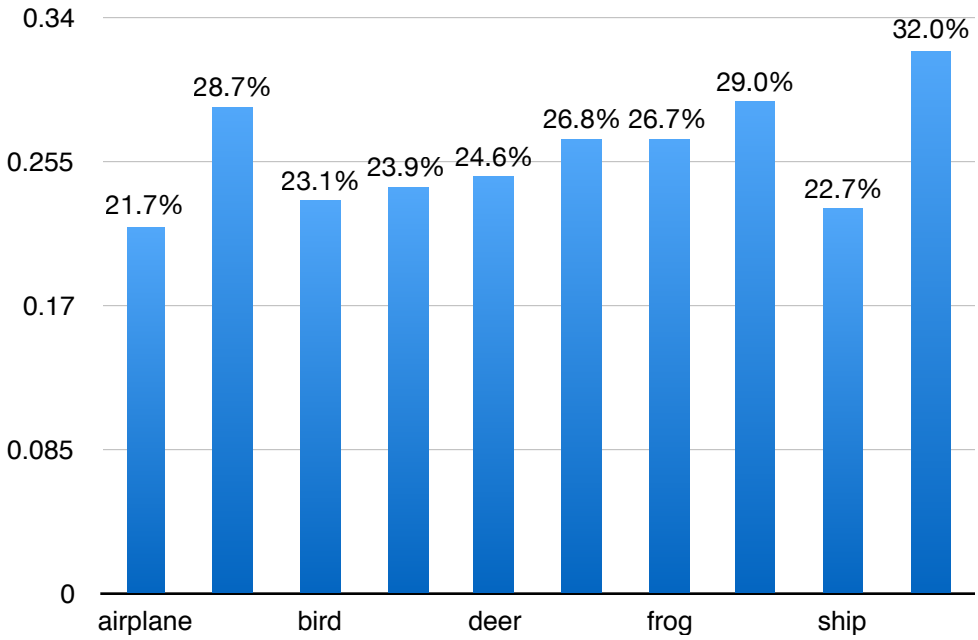
Problem 4.10

a. Running PCA on the 50k training images yielded the following mean images for each of the 10 categories:

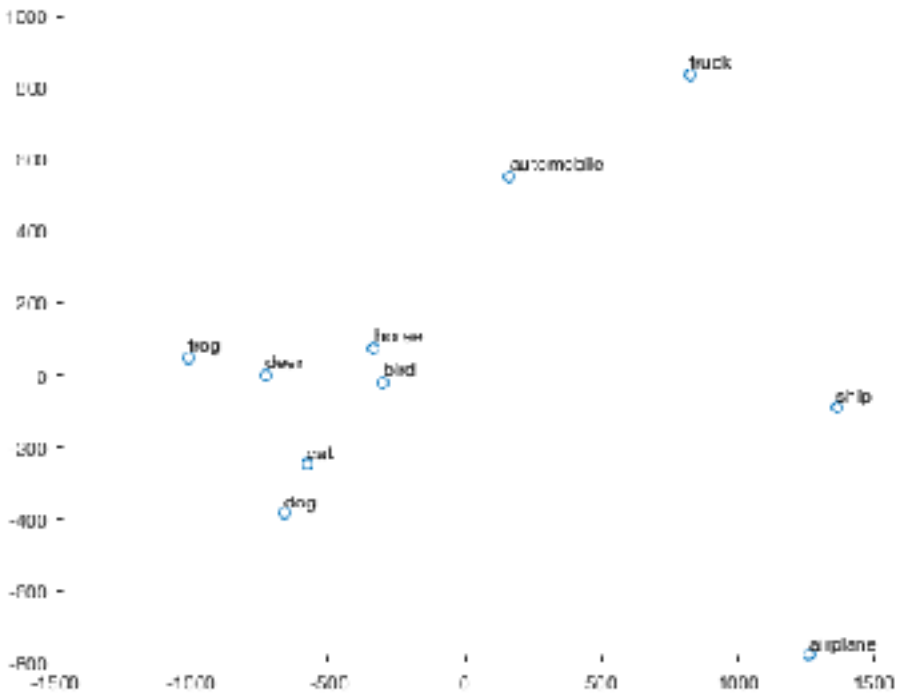


Note: these images have been rotated 90 degrees counter-clockwise.

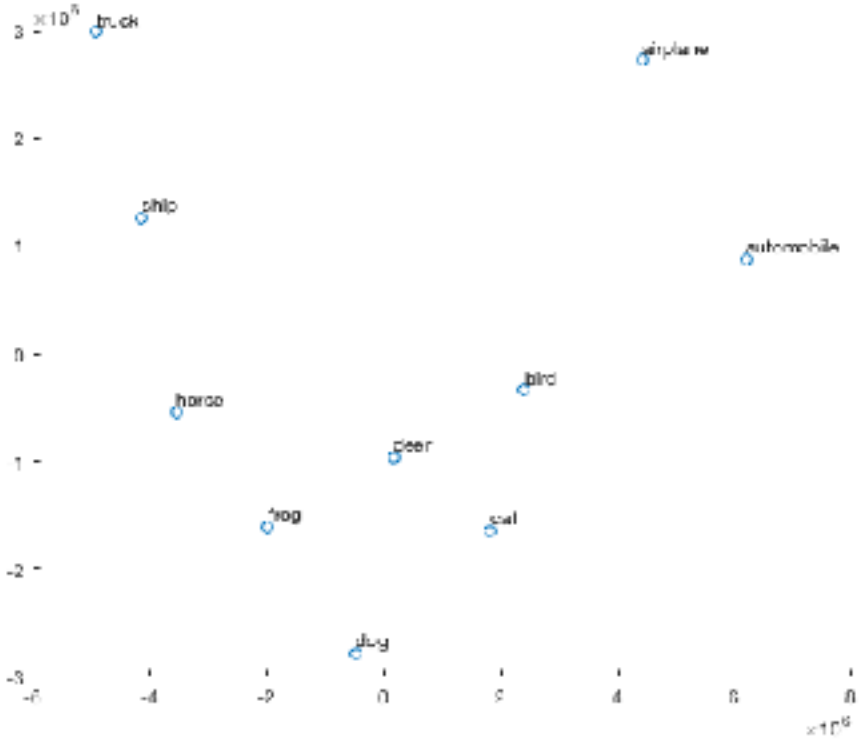
Additionally the average error resulting from representing the images of each category using the first 20 principal components was:



b. Treating the mean images as vectors and using distances between them for principal coordinate analysis yielded the following map of the means of each categories:



c. Defining the similarity between classes to be (E(A → B) + E(B → A))/2 and using this metric for principal coordinate analysis yielded the following 2D map of the classes:



The maps above are very different because they display different information. The map in part b shows the relative similarity between the average image of each category; for example: horse and bird have a very similar average image as seen by how closely they are positioned together, whereas truck and airplane have very dissimilar mean images seen by the large distance between those labels in the map.

The map in part c shows how well the first 20 principal components of each category can represent another. Using this map we can learn that you can represent a bird image using the first 20 principal components of deer as well as you can when the components of cat are used. Since the map in part c doesn't display the same information it makes sense that it doesn't look the same as the map in b.