hw3 - final-Copy5

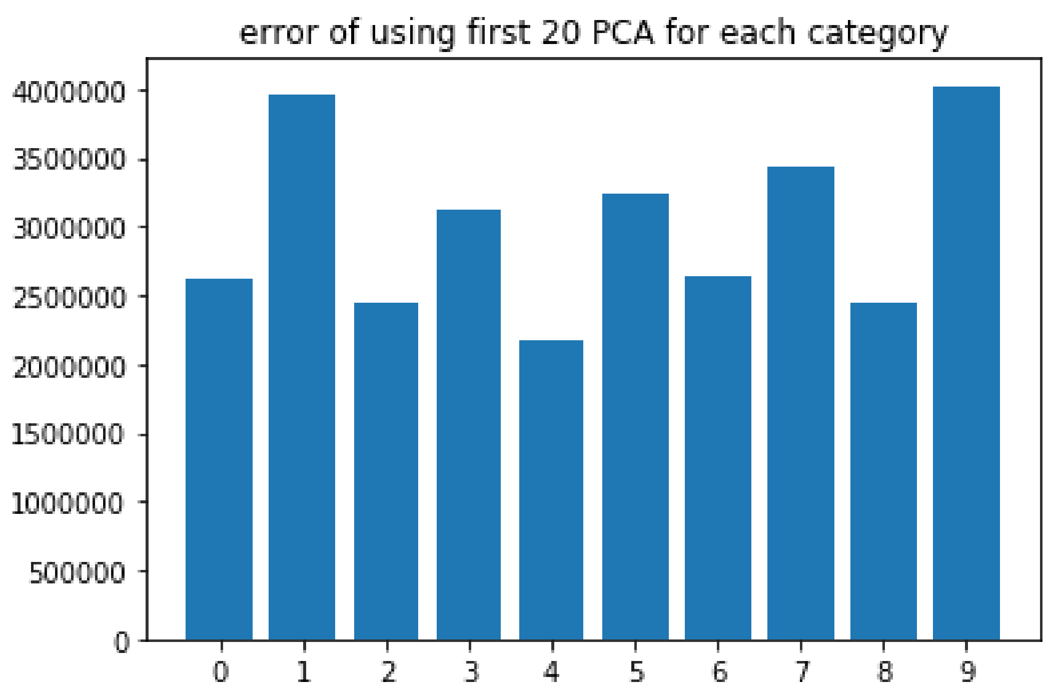
February 19, 2018

1 HW3

1.1 Q1

In [2]:

Out[2]: Text(0.5,1,'error of using first 20 PCA for each category')

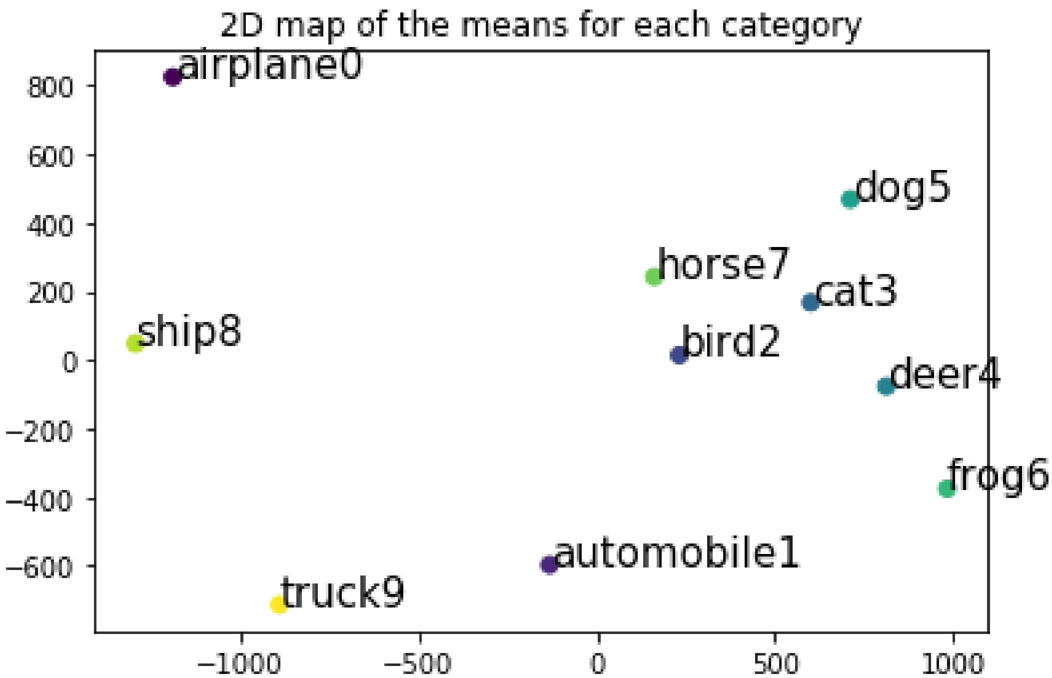


First I subtract the mean and put it into PCA, transform and inverse transform the data point mean image. This way I got the mean image using 20 principle at a low dimension. Use the sum of squared error and plot it for each category

1.2 Q2

Animals tend to cluster togerther while transportations is spaced out a little but still form a cluster which can easily be seperated from animals

In [11]:



Out[11]:

	0	1	2	3	4	5	6	7
0	0.00	1,683.64	1,605.02	1,905.54	2,148.76	1,965.22	2,445.68	1,663.65
1	1,683.64	0.00	886.24	1,027.65	1,143.08	1,216.08	1,191.19	950.79
2	1,605.02	886.24	0.00	517.31	601.25	701.47	913.75	418.28
3	1,905.54	1,027.65	517.31	0.00	469.79	412.18	677.49	596.38
4	2,148.76	1,143.08	601.25	469.79	0.00	617.70	460.51	684.35
5	1,965.22	1,216.08	701.47	412.18	617.70	0.00	828.58	843.67
6	2,445.68	1,191.19	913.75	677.49	460.51	828.58	0.00	948.70
7	1,663.65	950.79	418.28	596.38	684.35	843.67	948.70	0.00
8	945.54	1,303.47	1,557.72	1,851.21	2,065.62	1,897.59	2,249.20	1,660.27
9	1,449.09	950.00	1,416.67	1,676.47	1,830.74	1,880.24	1,913.24	1,347.33

	8	9
0	945.54	1,449.09
1	1,303.47	950.00
2	1,557.72	1,416.67
3	1,851.21	1,676.47
4	2,065.62	1,830.74
5	1,897.59	1,880.24

6	2,249.20	1,913.24
7	1,660.27	1,347.33
8	0.00	1,066.94
9	1,066.94	0.00

1.3 Q3

First I compute the pc on A by calling PCA,Then transform and inverse transform on B. This way I have the low dimenstion data. Compute the error and average them on the number of images

I test the diagonal errors using eigenvalues method and they match. the plot is similar as vehicles tend to cluster and animals form another cluster. They can be seperated by a line. The difference is the animals are not as close as in part 2. Using other category principle components to represent will cause more error than itself.This is also proved. Then compute similarity by the error will give the distance/dissimilarity matrix. The more error each pair compute, the more distant they are. Then call MDS and view it in a 2D space.

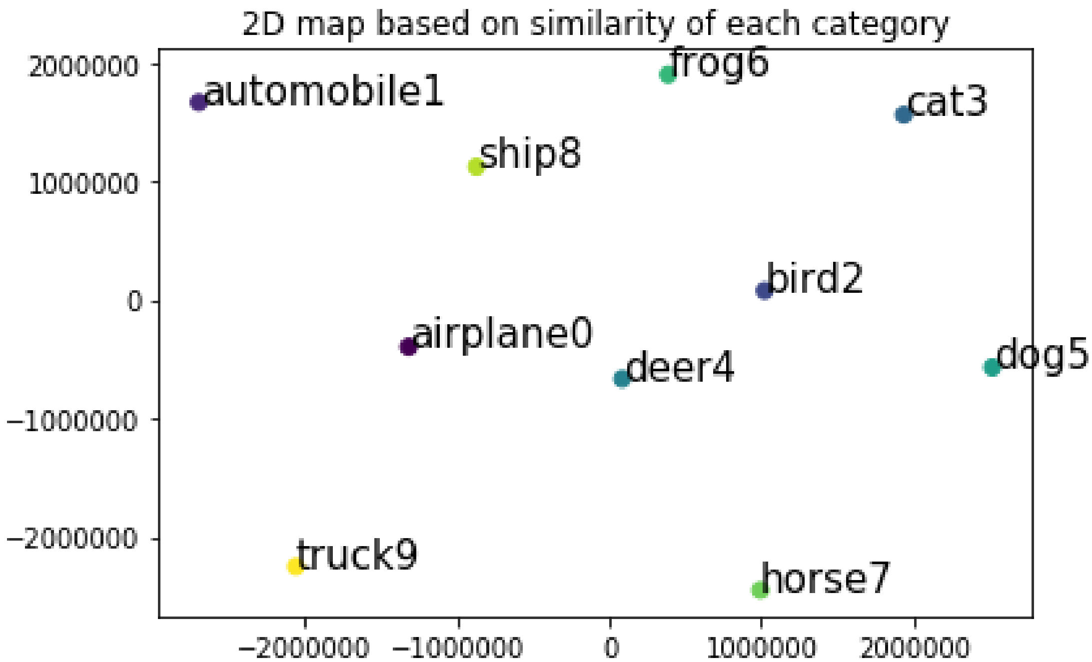
In [9]:

Out[9]:

	0	1	2	3	4 \
0	2,620,509.54	3,765,112.27	2,812,021.43	3,332,003.80	2,573,743.76
1	3,765,112.27	3,950,683.11	3,769,167.69	4,074,387.86	3,542,751.97
2	2,812,021.43	3,769,167.69	2,447,702.69	2,971,165.79	2,431,211.84
3	3,332,003.80	4,074,387.86	2,971,165.79	3,116,482.15	2,938,655.04
4	2,573,743.76	3,542,751.97	2,431,211.84	2,938,655.04	2,180,391.71
5	3,435,683.54	4,249,983.67	2,976,127.38	3,266,551.89	2,973,797.27
6	3,015,351.35	3,776,592.61	2,713,379.59	3,064,030.21	2,572,990.76
7	3,416,385.17	4,318,801.24	3,242,651.26	3,592,046.86	3,066,550.42
8	2,733,006.85	3,520,494.08	2,853,515.98	3,222,995.67	2,595,190.64
9	3,843,479.31	4,156,034.66	3,705,184.25	3,939,773.13	3,533,332.63

	5	6	7	8	9
0	3,435,683.54	3,015,351.35	3,416,385.17	2,733,006.85	3,843,479.31
1	4,249,983.67	3,776,592.61	4,318,801.24	3,520,494.08	4,156,034.66
2	2,976,127.38	2,713,379.59	3,242,651.26	2,853,515.98	3,705,184.25
3	3,266,551.89	3,064,030.21	3,592,046.86	3,222,995.67	3,939,773.13
4	2,973,797.27	2,572,990.76	3,066,550.42	2,595,190.64	3,533,332.63
5	3,231,171.35	3,127,057.91	3,647,693.44	3,421,118.51	4,124,617.44
6	3,127,057.91	2,630,248.61	3,411,787.98	2,935,355.98	3,751,892.73
7	3,647,693.44	3,411,787.98	3,441,096.05	3,433,660.90	4,214,646.94
8	3,421,118.51	2,935,355.98	3,433,660.90	2,440,637.17	3,592,892.89
9	4,124,617.44	3,751,892.73	4,214,646.94	3,592,892.89	4,021,107.26

In [9]:


2D map based on similarity of each category
automobile1
frog6
cat3
ship8
bird2
airplane0
deer4
dog5
truck9
horse7
2000000
1000000
0
−1000000
−2000000
−2000000
−1000000
0
1000000
2000000