

Homework 3
CS 498 - Applied Machine Learning
Larry Poon (lpoon2) & Titus Fong (*thfong2*)
25th February, 2018

Part 1

Error resulting from representing the images of each category using the first 20 principal components against the category.

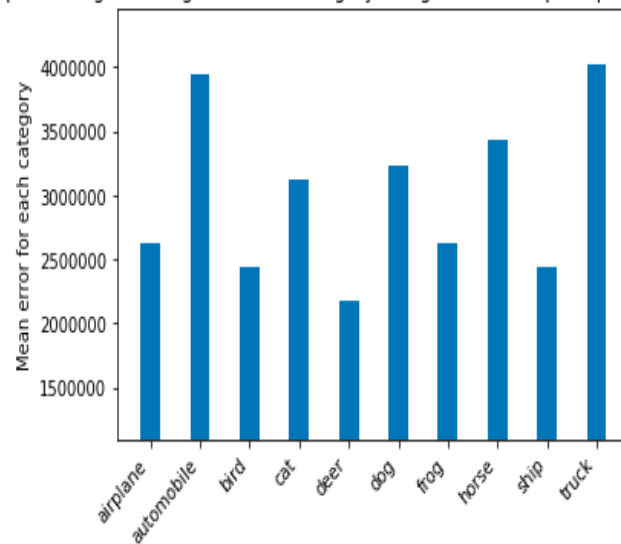


Figure 1. Error from representing images using first 20 components

Part 2

10x10 distance matrix

```
[[2.99009978e-04 1.68363535e+03 1.60502435e+03 1.90553526e+03
 2.14876341e+03 1.96522149e+03 2.44567973e+03 1.66364593e+03
 9.45541104e+02 1.44909491e+03]
[1.68363535e+03 0.00000000e+00 8.86236750e+02 1.02764978e+03
 1.14308137e+03 1.21607943e+03 1.19119200e+03 9.50786078e+02
 1.30346655e+03 9.49995771e+02]
[1.60502435e+03 8.86236750e+02 0.00000000e+00 5.17311502e+02
 6.01250335e+02 7.01468226e+02 9.13747516e+02 4.18276306e+02
 1.55771502e+03 1.41667473e+03]
[1.90553526e+03 1.02764978e+03 5.17311502e+02 0.00000000e+00
 4.69791716e+02 4.12181669e+02 6.77491969e+02 5.96376737e+02
 1.85121450e+03 1.67646786e+03]
[2.14876341e+03 1.14308137e+03 6.01250335e+02 4.69791716e+02
 0.00000000e+00 6.17697141e+02 4.60510929e+02 6.84346911e+02
 2.06562166e+03 1.83074085e+03]
[1.96522149e+03 1.21607943e+03 7.01468226e+02 4.12181669e+02
 6.17697141e+02 0.00000000e+00 8.28581051e+02 8.43672091e+02
 1.89759182e+03 1.88024377e+03]
[2.44567973e+03 1.19119200e+03 9.13747516e+02 6.77491969e+02
 4.60510929e+02 8.28581051e+02 0.00000000e+00 9.48704021e+02
 2.24919978e+03 1.91324088e+03]
[1.66364593e+03 9.50786078e+02 4.18276306e+02 5.96376737e+02
 6.84346911e+02 8.43672091e+02 9.48704021e+02 2.99009978e-04
 1.66026808e+03 1.34733410e+03]
[9.45541104e+02 1.30346655e+03 1.55771502e+03 1.85121450e+03
 2.06562166e+03 1.89759182e+03 2.24919978e+03 1.66026808e+03
 0.00000000e+00 1.06694163e+03]
[1.44909491e+03 9.49995771e+02 1.41667473e+03 1.67646786e+03
 1.83074085e+03 1.88024377e+03 1.91324088e+03 1.34733410e+03
 1.06694163e+03 0.00000000e+00]]
```

2D map of Euclidean distance of means of each categories

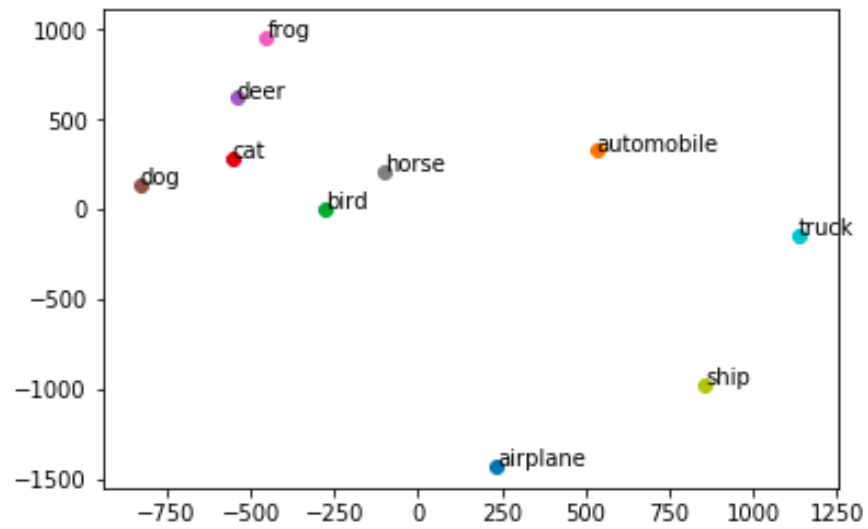


Figure 2. Euclidean distance between each class pair

Part 3

10 x 10 distance matrix

```
[[2620513.31910851 3765263.72482817 2812070.46621585 3332052.06860216
 2573709.08648721 3434991.39035774 3015350.62373824 3416927.28394457
 2733154.70995142 3843214.55891384]
[3765263.72482817 3950692.92519019 3769552.55317455 4074731.36436302
 3542961.47280071 4250212.61825153 3776667.02873286 4318665.24677164
 3520359.37697355 4155760.71610775]
[2812070.46621585 3769552.55317455 2447700.28710633 2970962.22771785
 2431224.06112306 2976043.11614289 2713257.04344166 3242350.08565994
 2853345.49003857 3705163.63526976]
[3332052.06860216 4074731.36436302 2970962.22771785 3116495.02927411
 2938744.15282851 3266247.79935133 3063927.31620294 3591865.72883422
 3222774.05864724 3939664.50671691]
[2573709.08648721 3542961.47280071 2431224.06112306 2938744.15282851
 2180394.80918182 2973998.746125 2572988.81061978 3066509.87728273
 2595121.88325287 3533484.13553322]
[3434991.39035774 4250212.61825153 2976043.11614289 3266247.79935133
 2973998.746125 3231131.71464675 3127194.38271375 3647466.21153927
 3420556.8672239 4124569.17970257]
[3015350.62373824 3776667.02873286 2713257.04344166 3063927.31620294
 2572988.81061978 3127194.38271375 2630248.5817849 3411669.76572176
 2935288.93298992 3752089.59822993]
[3416927.28394457 4318665.24677164 3242350.08565994 3591865.72883422
 3066509.87728273 3647466.21153927 3411669.76572176 3441110.2108379
 3433565.20681626 4214142.60555763]
[2733154.70995142 3520359.37697355 2853345.49003857 3222774.05864724
 2595121.88325287 3420556.8672239 2935288.93298992 3433565.20681626
 2440642.70790964 3592559.87115334]
[3843214.55891384 4155760.71610775 3705163.63526976 3939664.50671691
 3533484.13553322 4124569.17970257 3752089.59822993 4214142.60555763
 3592559.87115334 4021109.10789834]]
```

2D map of distances between each category by average error

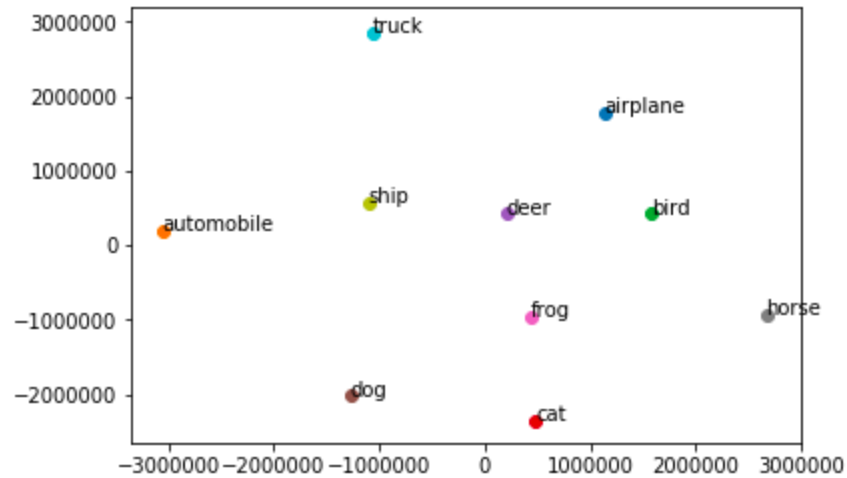


Figure 3. Measure similarity with different metrics

Comparison and analysis of figure 3 and 2:

First off, the two graphs are representing the measure of similarity between any pair of classes: figure 2 uses Euclidean distance between any 2 mean image classes, and figure 3 is measuring by average error by calculating the difference between each original images and the lower-dimensional representation of each image class. From what we see from the two graphs, figure 2 has more evident cluster, which means it's obvious to determine if one image class resembles or differs from others. However, data points in figure 3 seem like equally spread out, which means they are similarly difficult to be distinguished from other classes, unless the first 20 principle components of both comparing classes are somewhat similar (like bird and airplane class). Also, the distances measured by the method in figure 3 are significantly greater than those in figure 2. This can be explained as the basis (first 20 principle components) of every classes can poorly represent an image of a different class because each class image has its own selection of dimensions to be in the first 20 principle components, even the error between an image and the lower-dimensional representation using the basis of the same class is significantly large. Using Euclidean distance to measure similarity between classes seems to be a more direct and accurate measurement in this problem.