Assignment 3

CS498 Applied Machine Learning

Jianshu Wang - February 26, 2018

Introduction

Problem 1:

I get my mean figures by calculating mean values for all the vector , then I get for 10 mean figures for all the categories are in my folder, I list 3 of them below.

Figure 1: Mean Image of Bird, Airplane and Cat

These 3 mean images are quite fuzzy, and it represents the mean value of a class, I will use them for my calculation in problem 2.

In this problem I used sklearn PCA library for my calculation pca = PCA (n_components=20, svd_solver='full'), and then I get the errors vs categories figure which is shown below:

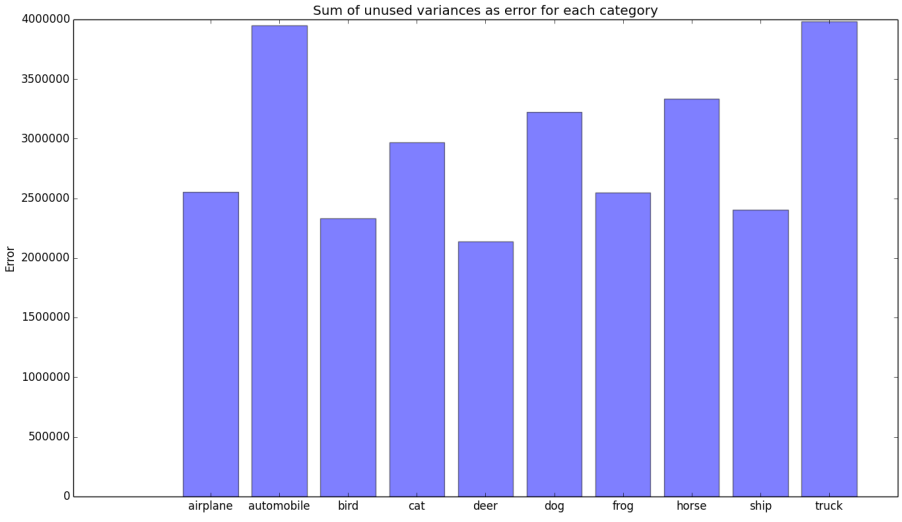

Figure 2: Error plot for Each Category

Problem 2:

This problem I used sklearn euclidean_distances library to calculate the euclidean distance matrix for 10 categories. Next I used manifold MDS libraries, here is code I use mds= manifold.MDS(n_components=2, dissimilarity="precomputed", random_state=None), in order to make it a 2D plot I set the n_component = 2, and none random_state. Then the result looks good from my knowledge, animal are in the same cluster and truck, ship and airplane are far away from which means they are not in the same group with these animals.

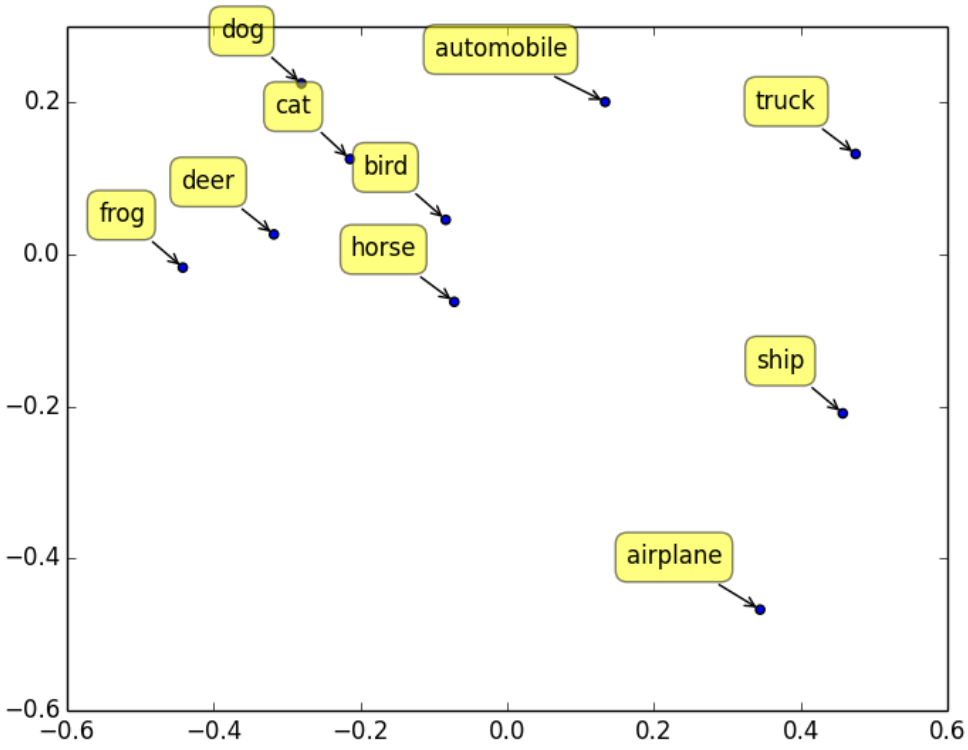

Figure 3: 2D Plot for Distance Matrix of Each Category

Here is the distance matrix:

[0. 1716.562 1634.185 1942.967 2175.085 2043.919 2528.438 1672.915 913.578 1460.198]
[1716.562 0. 879.660 988.844 1155.417 1232.328 1238.553 933.9444 1310.201 1067.424]
[1634.186 879.660 0. 547.389 615.700 744.324 976.240 458.223 1538.884 1512.629]
[1942.967 988.844 547.389 0. 492.409 419.861 699.142 661.838 1827.670 1760.344]
[2175.086 1155.417 615.700 492.409 0. 591.303 520.300 751.706 2065.039 1957.472]
[2043.919 1232.328 744.324 419.861 591.303 0. 821.903 927.496 1950.347 2029.705]
[2528.438 1238.553 976.240 699.1419 520.301 821.903 0. 1049.843 2299.285 2087.240]
[1672.915 933.944 458.223 661.838 751.706 927.496 1049.843 0. 1609.765 1396.175]
[913.578 1310.201 1538.884 1827.669 2065.039 1950.347 2299.285 1609.764 0 1053.439]
[1460.198 1067.424 1512.629 1760.344 1957.472 2029.705 2087.240 1396.175 1053.440 0.]

Problem 3:

In this problem, I used sklearn PCA library get all the pricipal components for each category, also I get xi as a big matrix for each category, what's more I kept their mean images for my following calculations.

$E(A|B) = actual(A) - [mean(A) + pc(B)]$, pc(B) is the part which contains the principal components. It is the sum of $[u_jT * (xi - mean(x))] * u_j$ of all the images in the same category. And the error should be the average value which is $(E(A|B) + E(B|A))/2$. Calculate the sum value of all items' square and take the mean, then we have a error value to fill a similarity matrix grid.

Then I used the same MDS library for the plot below

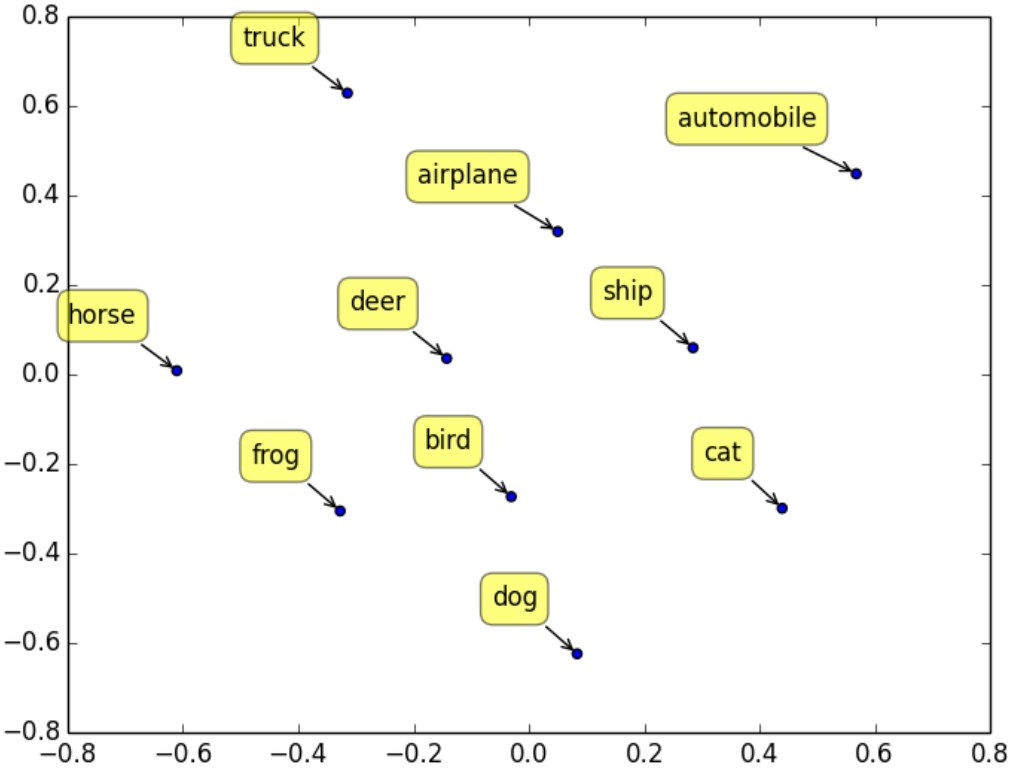

Figure 4: 2D Plot for Similarity Matrix of Each Category

[2550684 3805918 2892100 3371898 2662972 3574006 3033104 3473199 2793238 3939544]
[3805918 3947543 3802379 4154898 3580044 4414973 3785671 4359859 3606957 4381000]
[2892100 3802379 2331166 2932968 2471995 3100099 2711619 3244067 2881234 3761443]
[3371898 4154898 2932968 2968149 2913013 3331828 3065822 3570397 3259261 3976114]
[2662972 3580044 2471995 2913013 2134087 3102594 2631273 3076739 2630144 3590765]
[3574006 4414973 3100099 3331828 3102595 3223544 3226032 3713397 3523281 4208359]
[3033104 785671 2711619 3065822 2631273 3226032 2543918 3397799 2959211 3785749]
[3473199 4359859 3244068 3570397 3076739 3713397 3397799 3331221 3476051 4230338]
[2793238 3606957 2881234 3259261 2630144 3523281 2959212 3476051 2402619 3711554]
[3939543 4381001 3761443 3976114 3590765 4208359 3785749 4230338 3711554 3979199]