**MATH 6373: Homework 2**

Group: Carolina Garza, Michael Saenz, Ray Thomas

Question 0:

The dataset of this homework consists of 60,000 color images of 10 different classes of objects. The categories are the following: airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks. The dimensions of the images are 32 by 32 by 3. The data set was pre-split into a training set consisting of exactly 50,000 cases and a test set of 10,000 cases. A three layer MLP will be created with the goal of accurately classifying an unseen image.

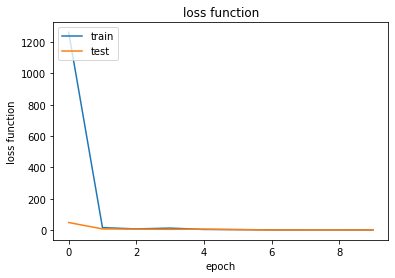
Question 1:

The MLP will consist of an input layer, a hidden layer, and an output layer with a softmax classifier to make the output more interpretable. The input layer has 3072 dimensions, the hidden layer initially has 1536 dimensions ( half the dimensions as the input layer), and the output layer has 10 dimensions due to there being 10 different classes. One hot encoding is used on response variables due to MLp’s not accepting categorical variables. The final output from the MLP will be a probability that the algorithm thinks that the input image belongs to a certain class.

The loss function we will use is CrossEntropy. Cross entropy is calculated by taking the negative log of all the outputs from the softmax function. The average cross entropy among all the classes will be used to track the performance of the MLP.

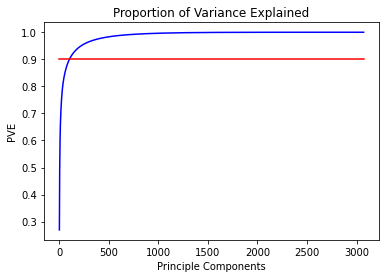
The first 5 response values from the response variable are checked to make sure they align with the 5 values after encoding, which they do. The length of the features total 3072. The data is flattened to reduce the 32x32x3 to one dimension with the length 3072.

At this stage, the initial MLP is constructed using batches of 1000 and 10 epochs. The loss function results after each epoch are graphed below.



Question 2:

In this step, PCA will be conducted to find the smallest h such that 90% of the variance is retained. Automatic learning will be launched with a batch size of 100, 50 epochs, and cross entropy as the loss function. To do this, the data is scaled and the eigenvectors and eigenvalues are extracted. Iterating cumulatively through the eigenvalues yields that 103 is the minimum number of eigenvalues needed to reach 90% of the explained variance.

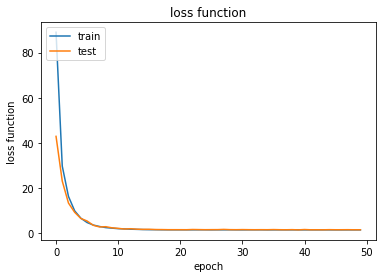


The next step is to find the minimum number of epochs such that the average cross entropy across epochs stabilizes for the training set, the average cross entropy for the test set reaches and global minimum and starts to rise from there, and that training entropy is less than testing entropy. This is to prevent overfitting to the training dataset. The optimal number of m epochs is found to be 22. Using m = 22, the following confusion matrices are generated on the training and test data sets.

Confusion matrices

| Training set confusion matrix | Test set confusion matrix |
| --- | --- |
|  |  |

By looking at the diagonal of each matrix, the matrices have mostly better than 50% accuracy and this is acceptable.



Question 3:

For each of the 10 classes, PCA is conducted for all of the input belonging to each respective class. In every class, most of the variance is captured well before 500 components are used. This is significant since the number of components needed can be drastically reduced with this PCA.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

All classes combined

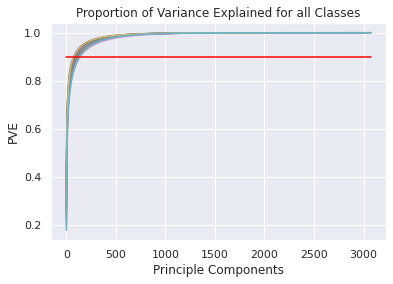


Table of the number of necessary components needed to reach 90% of the explained variance

| h1 | h2 | h3 | h4 | h5 | h6 | h7 | h8 | h9 | h10 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 75 | 123 | 87 | 83 | 103 | 91 | 136 | 116 | 77 | 137 |

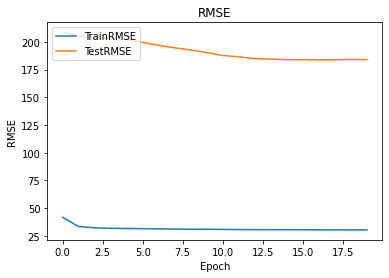
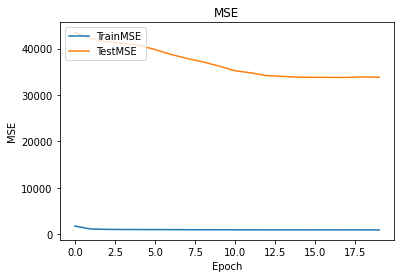
Confusion matrices

| Training set confusion matrix | Test set confusion matrix |
| --- | --- |
|  |  |

Question 4:

The goal of this step is to improve the MLP given this new information from the previous steps. The new H is defined as the sum of the h’s in the previous table, which equals 1028. This will be the new h for the new and improved MLP.

To move forward, an autoencoder is necessary. To do this, an h2 needs to be found such that 95% of the variance can be explained. After doing PCA, h2 is found to be 752. h2 will be used as the h for the encoder, and the new, high h is 1028 for the decoder. The encoder is now trained using batches of 100. The results are in the graph below.



Question 5:

Using the weights and thresholds calculated in the first two layers of the encoder, a new MLP with the same weights and thresholds for layers 1 and 2. The loss function still remains cross entropy.

Question 6:

Finally, the complete long MLP has results: