|  |
| --- |
| **Report** |

1. **CNN Pattern Classifier**

For solving the problem given in the assignment, I used simple two layers CNN model with 4 fully connected layers as a classifier in the both parts of the question. The CNN model, training process, and testing process are all initially copied from this website and then edited later <https://adventuresinmachinelearning.com/convolutional-neural-networks-tutorial-in-pytorch/.>

1. **Solving Mix Train Data Set Method**

My approach of solving the mixed training data set is by cleaning the data set. The first way of cleaning the data set is calculating, for every image, the difference from all other images whose share the same class with. After that, I calculate the standard deviation and delete every image that is less than minus 1.5 of the standard deviation or more than 1.5 of the standard deviation. By doing so, I’m able to delete every image that has high difference from other images of its same class. This means that if the image is very different then all the other images of its class then it’s in the wrong class. The problem of this method is that it’s hard to compare two arrays together whether they are greater than each other or not. So, I tried to take the mean of the standard deviation array, which I believe is a weak approach, and the mean of the image itself and compared them together. Since calculating the mean is not a perfect method and may generate other problems, I tried to add another method to cope with the mean mistake since comparing two arrays is hard. The second approach of cleaning the data is by using a simple classifier and reclassify the whole database. In my case, I have used the Knn classifier, with n equals 10, since it classifies the image with nearest image which seem perfect to our problem. However, there are many problems that can rise from this approach. One of the problems are what if the classifier classified a correct image wrongly! In this case, I just combine the old data set with the new classified data set so that we won’t lose any valuable information. By doing so, we can actually substitute the data set from the loss it suffered in the cleaning data step, first method.

1. **Results and Accuracy**

I tested the CNN classifiers three times once on the well organized data set and twice on the mix training data set, first without applying just registering the model behavior and the second using the clean and reclassify data set approach. The cleaned reclassified data set did very well in the training phase even better than the not cleaned mixed data set but poorly in the testing phase. I believe the poor results obtained in the testing phase can be solved by changing the hyper parameters like the n value in the KNN classifier and the condition that the mean can be less or greater than 1 standard deviation or 2 standard deviation and so.

This table shows used parameter in all training process.

|  |  |
| --- | --- |
| num\_epochs | 2 |
| batch\_size | 25 |
| learning\_rate | 0.001 |

This table shows accuracy obtained on the test data set.

|  |  |
| --- | --- |
| Test accuracy of model trained on the well organized data set | 98.5% |
| Test accuracy of model trained on the mixed data set | 95.45% |
| Test accuracy of model trained on the mixed data set using clean and reclassify data set approach | 30.62% |

These images shows the training process for all the three cases.

|  |  |
| --- | --- |
| model training on the well organized data set | firstfirst2 |
| model training on the mixed data set | secondsecond2 |
| model training on the mixed data set using clean and reclassify data set approach | thirdthird2 |

|  |  |
| --- | --- |
| Student Name: | Safaa Moallim |
| Student ID: | 2019227144 |