# Report :: TIPR Assignment - III

A. Lakshmi

4 Mar, 2019

## 1 part I: CIFAR-10

The following section gives the details of the experiment conducted on CIFAR-10. The accuracy for training data is around 96 %, whereas the accuracy for test data is very less around 60 %. the architecture had varying number of convolutional layers, each convolutional layer followed by max pool layer and batch normalization layer. The convolutional layer was followed by 2 fully connected layers with 128 and 64 hidden neurons. The final fully connected layer has 10 neurons corresponding to 10 classes. When the number of fully connected layers increased, the performance of the classification deteriorated.

### 1.1 Task I: Number of layers

Fig. 1 shows the plot of accuracy, F1 macro and micro vs number of layers. The number of convolutional layers were varied from 3 to 7. There was no change in the number of fully connected layer. The filter size was held constant. It could be inferred that the network provides good accuracy for lesser number of convolutional layers.

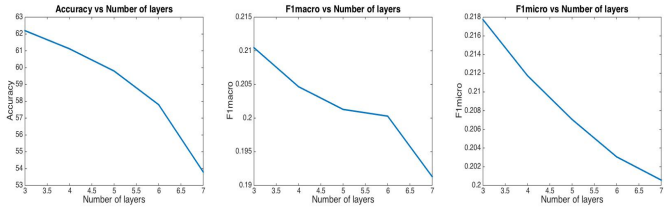


Figure 1: Plot of accuracy, F1 macro and F1 micro vs number of layers

### 1.2 Task 2: Size of the Filter Kernel

Fig. 2 shows the plot of accuracy, F1 macro and micro vs size of the filter kernel. In this case, the number of convolutional layers were kept constant as 3. The size of the filter kernel was varied from 3 to 7. It could be inferred that

the performance is better for lesser size of the filter as it captures more local information in the images.

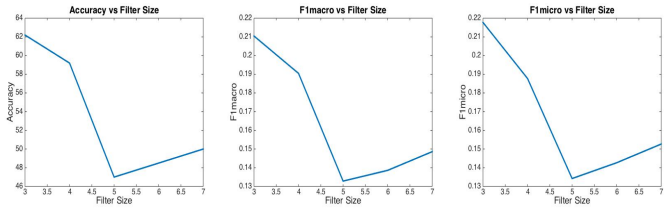


Figure 2: Plot of accuracy, F1 macro and F1 micro vs Filter size

## 1.3 Task 3: Activation functions

Fig. 3 shows the plot of accuracy, F1 macro and micro vs activation functions. The number of layers (both convolutional and fully connected) and other parameters of the network were held constant. The different activation functions experimented were ReLu, sigmoid, tanh and Swish. It could be inferred from the figure that the performance is better for swish, tanh and ReLu, best perfromance being achieved for swish. The performance deteriorates for sigmoid activation.

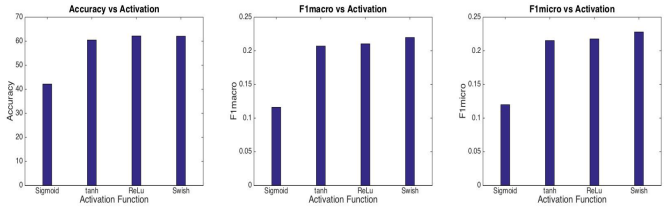


Figure 3: Plot of accuracy, F1 macro and F1 micro vs Activation function

## 1.4 Task 4: Initalization

Fig. 4 shows the plot of accuracy, F1 macro and micro vs initializations. In this case, all the other parameters of the network were held constant. The activation function used is ReLu. The different initialization tried were trncated normal, normal uniform distributions and constant value. It could be inferred from the figure that the performance is better for truncated normal, normal and uniform distributions, the best being achieved from normal distribution.

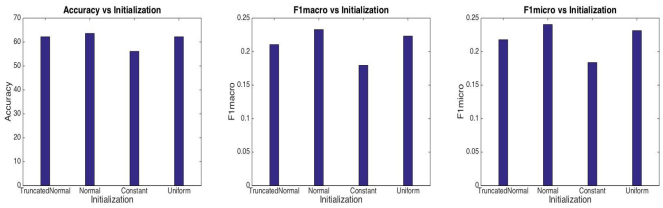


Figure 4: Plot of accuracy, F1 macro and F1 micro vs Initialization

## 1.5 Task 5: Clustering

Fig. 5 shows the plot of maximum clustering accuracy as mentioned in the assignment vs the ratio of data considered for training. 10 to 50 % of the total data was used to train the network. The embedding of the penulitmate layer was extracted for the remaining test data. The embeddings are subjected to K-Means clustering algorithm. As K-Means will assign clusters labels randomly which may not correspond to the original cluster label, the accuracy was estimated for different permutations of K-Means cluster label and the best accuracy is considered as the true accuracy.

It could be inferred from the figure that the best accuracy if obtained when more data is used for training. With more data, the network learns the pattern in the data with more accuracy and the embeddings of the penultimate layer becomes better representation of the data.

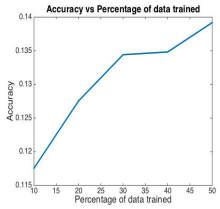


Figure 5: Plot of accuracy of KMEANS clustering algorithm vs the number of training data points

## 1.6 Task 6: t-SNE visualization

Fig. 6 shows the 2 dimensional clustered plot of t-SNE visualization. The embeddings of the penultimate layer is provided to TSNE function of python which finds two dimensional representation of the embedding. It could be inferred from the figure that the network has not learnt enough to represent the data as well defined clusters.

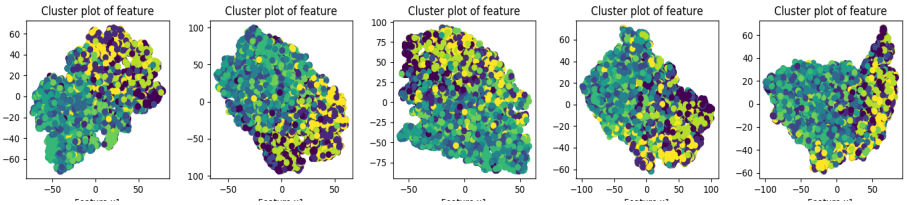


Figure 6: Clustered output of t-SNE visualization

### 1.7 Task 7:Comparison with MLP

Fig. 7 shows the comparison plot of MLP and CNN in terms of accuracy, F1 macro and micro vs number of layers. The number of hidden layers and number of convolutional layers were varied in MLP and CNN respectively. The filter size was fixed to 3. The activation function used was ReLu. It could be seen from the figure that CNN outperforms MLP as CNN captures structural information as well

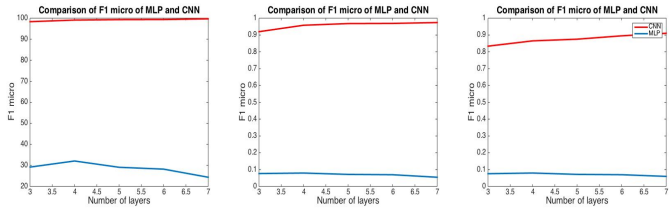


Figure 7: Plot of comparison of MLP and CNN vs number of layers in terms of accuracy, F1 macro and F1 micro vs Initialization

## 2 part 2: Fashion-MNIST

The following section gives the details of the experiment conducted on Fashion-MNIST. The accuracy for test data is around 98 %, as against the accuracy that was achieved with CIFAR-10. The architecture had varying number of convolutional layers, each convolutional layer followed by max pool layer and batch normalization layer. The convolutional layer was followed by 4 fully connected layers with 128, 256, 512 and 1024 hidden neurons. The final fully connected layer has 10 neurons corresponding to 10 classes.

### 2.1 Task I: Number of layers

Fig. 8 shows the plot of accuracy, F1 macro and micro vs number of layers. The number of convolutional layers were varied from 3 to 7. There was no change in

the number of fully connected layer. The filter size was held constant. It could be inferred that the network provides good accuracy with increase in number of convolutional layers.

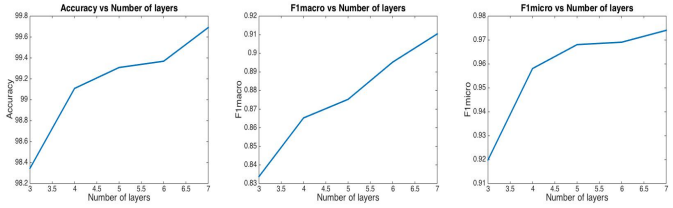


Figure 8: Plot of accuracy, F1 macro and F1 micro vs number of layers

## 2.2 Task 2: Size of the Filter Kernel

Fig. 9 shows the plot of accuracy, F1 macro and micro vs size of the filter kernel. In this case, the number of convolutional layers were kept constant as 3. The size of the filter kernel was varied from 3 to 7. It could be inferred that the performance is better for certain sizes of filter kernels.

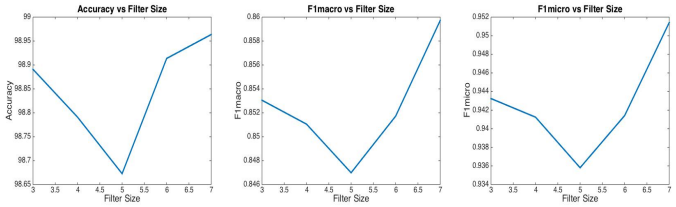


Figure 9: Plot of accuracy, F1 macro and F1 micro vs filter size

## 2.3 Task 3: Activation functions

Fig. 10 shows the plot of accuracy, F1 macro and micro vs activation functions. The number of layers (both convolutional and fully connected) and other parameters of the network were held constant. The different activation functions experimented were ReLu, sigmoid, tanh and Swish. It could be inferred from the figure that the performance is better for swish, tanh and ReLu, best perfromance being achieved for swish. The performance deteriorates for sigmoid activation.

## 2.4 Task 4: Initalization

Fig. 11 shows the plot of accuracy, F1 macro and micro vs initializations. In this case, all the other parameters of the network were held constant. The activation

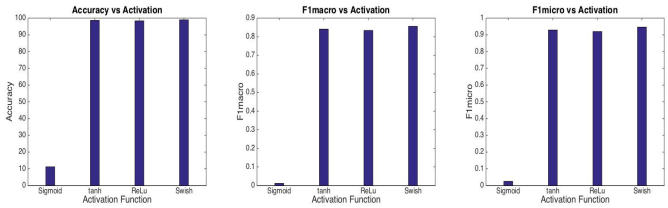


Figure 10: Plot of accuracy, F1 macro and F1 micro vs Activation function

function used is ReLu. The different initialization tried were truncated normal, normal uniform distributions and constant value. It could be inferred from the figure that the performance is better for truncated normal, normal and uniform distributions, the best being achieved from normal distribution.

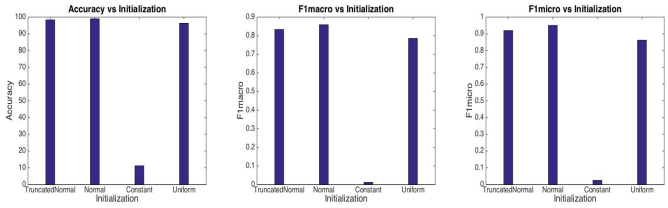


Figure 11: Plot of accuracy, F1 macro and F1 micro vs Initialization

## 2.5 Task 5: Clustering

Fig. 12 shows the plot of maximum clustering accuracy as mentioned in the assignment vs the ratio of data considered for training. 10 to 50 % of the total data was used to train the network. The embedding of the penulitmate layer was extracted for the remaining test data. The embeddings are subjected to K-Means clustering algorithm. As K-Means will assign clusters labels randomly which may not correspond to the original cluster label, the accuracy was estimated for different permutations of K-Means cluster label and the best accuracy is considered as the true accuracy.

It could be inferred from the figure that the best accuracy is obtained when more data is used for training. With more data, the network learns the pattern in the data with more accuracy and the embeddings of the penultimate layer becomes better representation of the data.

## 2.6 Task 6: t-SNE visualization

Fig. 13 shows the 2 dimensional clustered plot of t-SNE visualization. The embeddings of the penultimate layer is provided to TSNE function of python

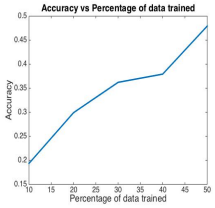


Figure 12: Plot of accuracy of KMEANS clustering algorithm vs the number of training data points

which finds two dimensional representation of the embedding. It could be inferred from the figure that the network has learnt enough to represent the data as well defined 10 clusters.

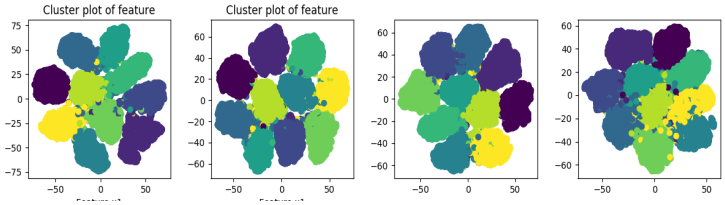


Figure 13: Clustered output of t-SNE visualization

## 2.7 Task 7:Comparison with MLP

Fig. 14 shows the comparison plot of MLP and CNN in terms of accuracy, F1 macro and micro vs number of layers. The number of hidden layers and number of convolutional layers were varied in MLP and CNN respectively. The filter size was fixed to 3. The activation function used was ReLu. It could be seen from the figure that CNN outperforms MLP as CNN captures structural information as well

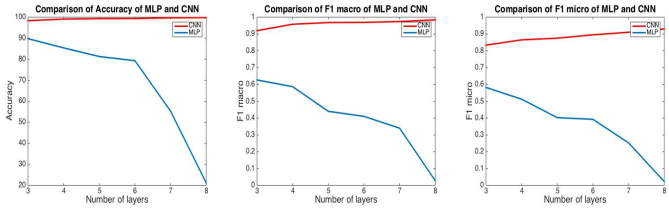


Figure 14: Plot of comparison of MLP and CNN vs number of layers in terms of accuracy, F1 macro and F1 micro vs Initialization