ECE 657 ASSIGNMENT THREE SPRING 2021 OKPALA WALTER ID:20910035
HALIM RATMONO ID: 20921090
SAMARTH JAIN ID

QUESTION ONE REPORT

PREPROCESSING STEPS

Normalization

The dataset was first converted from integer to float before the normalization process was carried out on it and this is done to avoid rounding problems during the normalization. The normalization process is carried out to format the dataset in such a way that the mean is close to zero, this is done to speeds up learning which leads to faster convergence.

Reshape

Another preprocessing step that was taken was to reshape the dataset from (50000 , 32 , 32 , 3) to

(50000, 3072) so that it can match the format for the fully connected neural network(MLP) input layer.

DESCRIPTION OF THE OUTPUT LAYER USED

The output layer used has 10 unites given the fact that the number of classes are ten. The activation function employed in this layer is the Softmax activation function and this is done because the Softmax function takes the logit produced by the previous layer of the network and the then outputs a probability distribution over the classes which is what is required to obtain the class which the given input image belongs to.

DESCRIPTION OF LOSS FUNCTION

The loss function employed for this classification problem is Sparse Categorical Cross entropy, this loss function was used because

1 The labels where integer and not one-hot encoded

2 The classes are mutually exclusive (I.e., each sample in the dataset belongs to only one class)

ECE 657 ASSIGNMENT THREE SPRING 2021 OKPALA WALTER ID:20910035
HALIM RATMONO ID: 20921090
SAMARTH JAIN ID

QUESTION ONE REPORT

TRAINING AND VALIDATION ACCURACY FOR VARIOUS MLP NETWORKS

models	training accuracies	validation accuracies	num_hidden_layer	num_of_parameters
1.0	0.507	0.487	2.0	1841162.0
2.0	0.499	0.474	3.0	2103818.0
3.0	0.484	0.478	4.0	2366474.0
4.0	0.467	0.451	5.0	2629130.0
5.0	0.45	0.438	6.0	2891786.0
6.0	0.444	0.428	2.0	206282.0
7.0	0.432	0.419	3.0	276170.0
8.0	0.474	0.453	4.0	1746506.0
9.0	0.386	0.388	5.0	436554.0
10.0	0.353	0.358	6.0	1748634.0

From the table above we discover that the first MLP network which is made up of 1,841,162 parameters and two hidden layers has the highest training and validation accuracy as compared to other models with higher depth(number of hidden layers) and larger amount of parameters also network with large depth and less parameters tend to perform badly (e.g., model 9), so we can infer from the table that the higher depth of the network in this case reduce the training and validation accuracy while the number of higher number parameters tend to increase the training and validation accuracy.

TRAIN AND TEST ACCURACY FOR ALL THREE NETWORKS:

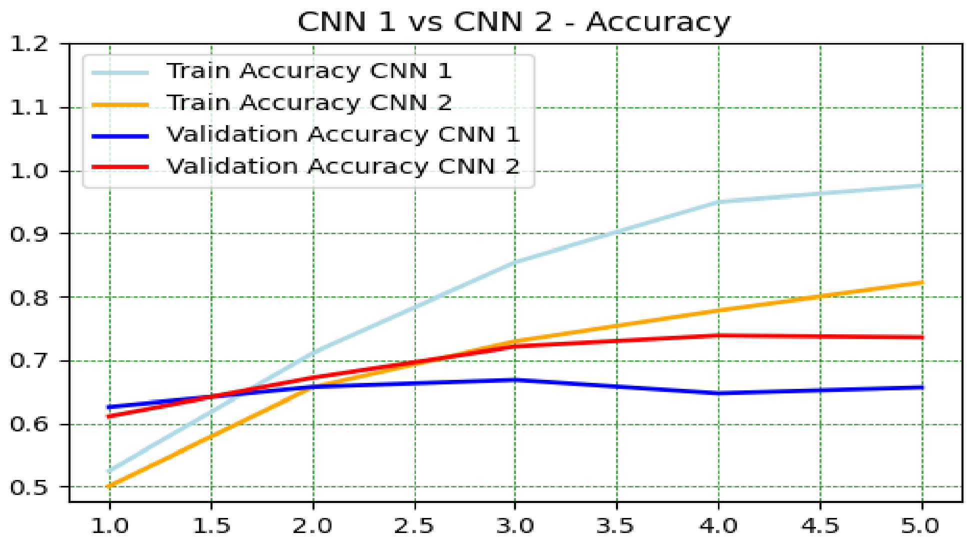
	training accuracies	test accuracies
MLP	0.508	0.481
CNN1	0.975	0.645
CNN2	0.822	0.729

From the above table we can deduce that both training and testing accuracy for the CNN models are very high as compared to the MLP models and the reason is. Fully connected neural networks[FCNN] don’t see any order in their inputs , this implies that if inputs are shuffled , FCNN will have the same training and test accuracy it had when it was not shuffled, this implies that FCNN does not take into account the spatial relationship between the pixels of the image. CNN on the other hand considers the spatial relationship between pixels which are close together (i.e., local spatial coherence of images) , first this is good because the number of operations to learn the image is reduced as convolutions on patches of adjacent pixels is employed , this property of CNN tends to increase its performance.

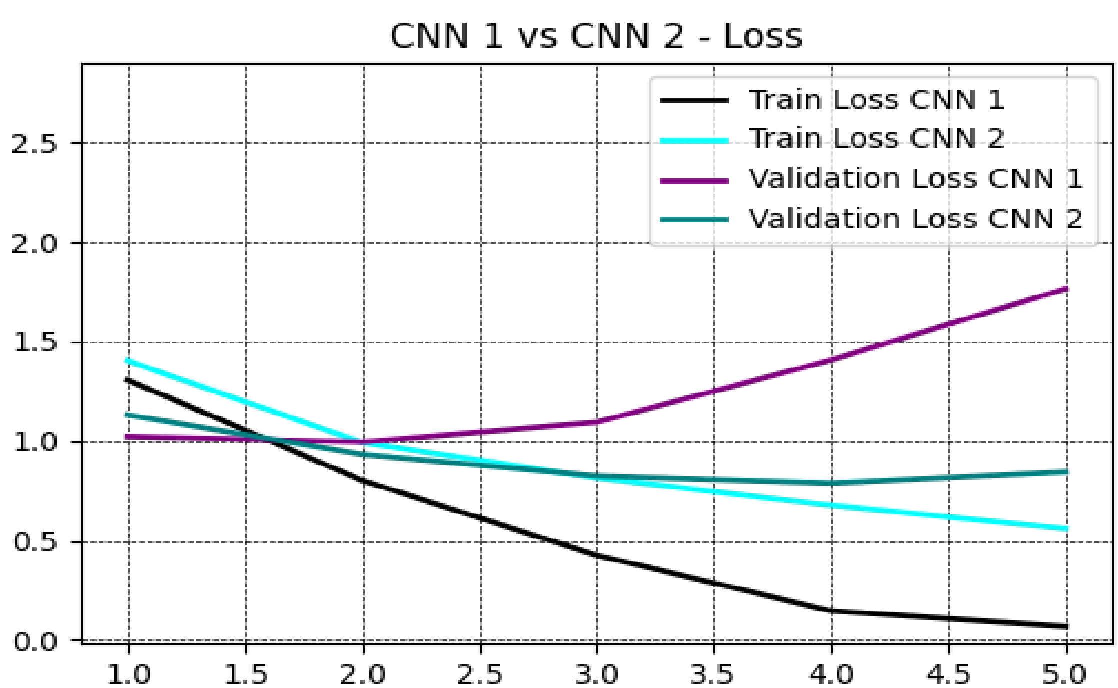
ECE 657 ASSIGNMENT THREE SPRING 2021 OKPALA WALTER ID:20910035
HALIM RATMONO ID: 20921090
SAMARTH JAIN ID

QUESTION ONE REPORT

TRAINING AND VALIDATION ACCURACY CURVES FOR THE TWO CNNS



TRAINING AND VALIDATION LOSS CURVES FOR THE TWO CNNS



ECE 657 ASSIGNMENT THREE SPRING 2021 OKPALA WALTER ID:20910035
HALIM RATMONO ID: 20921090
SAMARTH JAIN ID

QUESTION ONE REPORT

From the above graphs we can deduce that the first model has a high training accuracy and a low training loss as compared to the second model, but on the other hand the validation accuracy for the second model is grater than that of the first model and so is the validation loss of the second model less than that of the first model. This is observation due to the dropout layer introduced in the second model.

COMPARING TRAINING TIME FOR THE CNN MODELS

Total trained time for the first CNN Is 46.60 seconds

Total training time for the second CNN is 85.85 seconds

From the above data we can deduce that the second CNN model take more time to train than the first CNN model. This could be attributed to the addition of pooling layer and dropout layer operation in the second model.

RELATING CNN ARCHITECTURS TO RESULTS

From the graphs for training and validation accuracy of both CNN models we can deduce that the training accuracy for the first CNN is grater than that of the second but for the validation accuracy for the second is grater than of the first and this due to the dropout layer(**Regularization**) introduced in the second CNN model this layer increase the bias error (which reduces the training accuracy and prevents overfitting) but reduces the variance error which makes the network to better generalize on the unseen dataset(validation dataset) as compared to the first CNN model.

EXPECTATION OF ACCURACY IF NUMBER OF EPOCHS IS INCREASED

The accuracies of the models will increase if the number of epoch where increase because increasing the number of epochs will increase the number of times gradient decent operation is carried out and this in turn reduces the loss and increase the accuracy, but care should be taken to prevent overfitting by finding that sweet spot for the number of epochs which produces an optimal training , validation, and test accuracy of the models.

RECOMMENDATIONS TO IMPROVE THE NETWORK.

One of the ways to improve the network is to add more convolutional layers, this more convolutional layer that is added to the more features of the image the network can extract and this leads to better representation of the image and hence increase accuracy. Care should be taken when increasing the number of convolution layer because at a point(or a certain threshold) instead of the added layer to extract features from its input it tends to overfit the data, and this leads to issue like false positives during prediction.