Visual Recognition

Assignment 3

Ishaan Sachdeva

IMT2018508

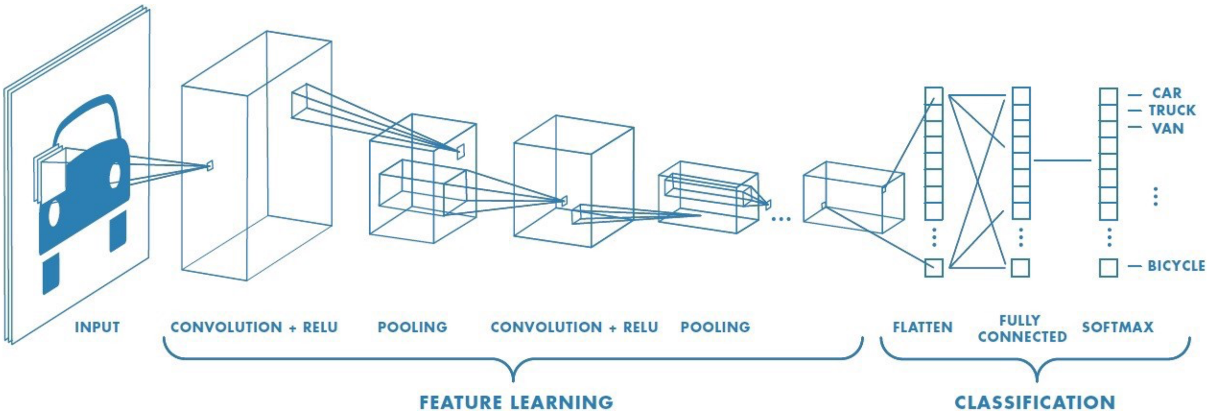
Task1

In this task, we have to implement the Alexnet convolutional neural network on the cifar-10 dataset.

Artificial Neural Networks(ANN) are non linear models with group of neurons at each layer. It consists of 3 layers: input, hidden, output layer. The input layer accepts the input , the hidden layer processes the input and the output layer produces the result. Since inputs are processed only in the forward direction they are also called as feed forward neural network.

ANN are capable of learning any non linear model and hence they show better accuracy as compared to the linear models but incase of images, an 2d image needs to be converted to 1d array to train a model. for eg an 256*256 pixel image, the number of trainable parameters will be 2,62,144 for 1 hidden layer with 4 parameters.

CNN(Alexnet)



CNN is a deep learning algorithm which takes image as input and assigns(importance/weights) to various aspects/objects in a region. The architecture is analogous to the connectivity pattern of human neoron where individual neuron respond to to stimuli in a specific region of visual field.

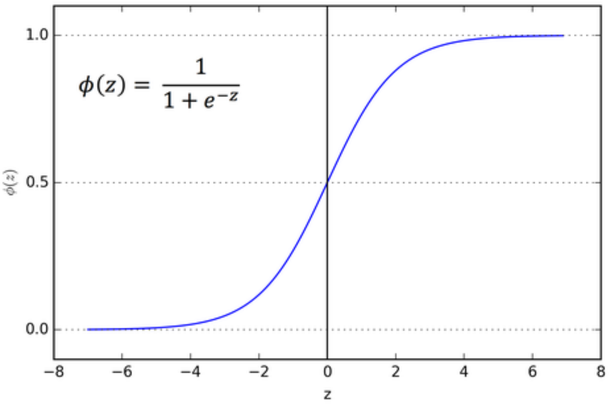
In feature engineering of CNN, **convolution operation** is done on the image which extract features from a given image through relevant filters. The best part about this operation in a CNN network is it learns the filters automatically and nothing needs to specified explicitly. As discussed in ANN, trainable weights in case of NN are a lot and so **Pooling** is the method used to reduce the size. This reduces the computational power required to process the data through dimensionality reduction. It is also used in extracting features which are rotational and position invariant features. After the features are extracted, it is converted to a 1d array(**Flatten layer**) and given to the input layer of an ANN model with input, hidden(**Dense layer**), output layer for training the model. The input size at each layer and trainable parameters can be viewed using the .summary() function.

Activation Function

Activation functions are used to determine the output of neural network like 0 or 1. They are the functions which introduce non linearity to the network. There are different kind of activation functions:

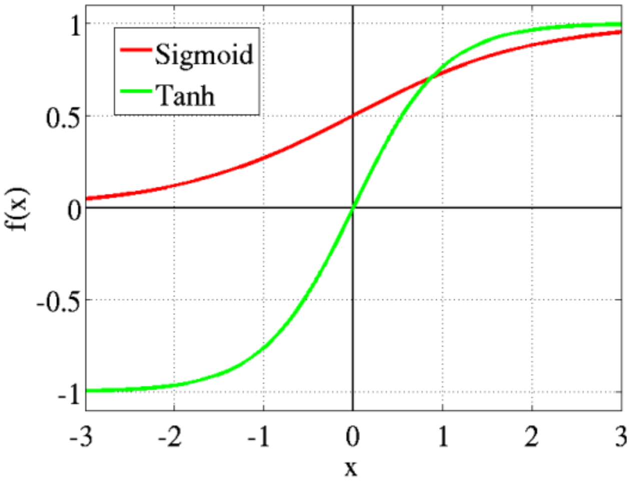
1) Sigmoid

The output of Sigmoid function lies between 0 and 1. It is used in places where we need to predict the probability.



2) tanh

tanh is similar to sigmoid function instead it lies in the range -1 to 1.

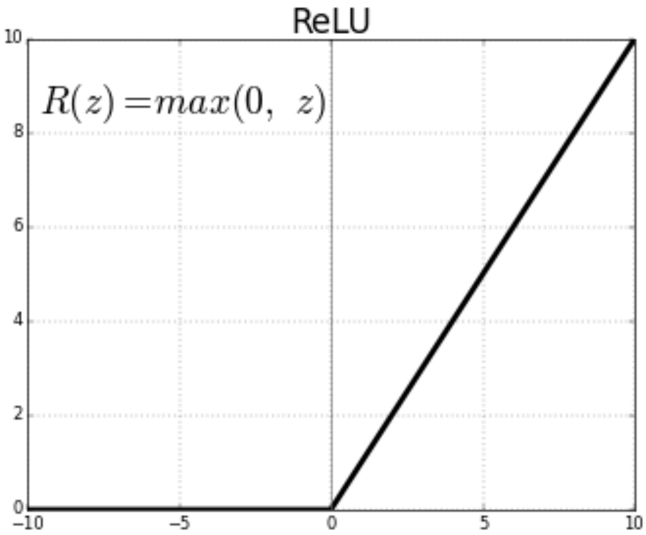


The issue with both tanh and sigmoid function is the vanishing gradient problem. While training a neural network we need to calculate the gradient descent and back propagate the errors but in the case of sigmoid and tanh at very high and low values of x the gradient is close to zero which slows the learning process and so we use ReLU.

3) ReLU

It is the most commonly used activation function **.**

It is defined as max(0,x)



Training the model

The dataset used is the cifar-10 dataset which has 60000 images divided into 10 classes. 50000 images are used to train the model and 10000 images are used to test the model.

The CNN network used has 3 convolutional layers and 2 Fully connected layers.

Below table represents the accuracies of different model.

MODEL	ACCURACY	Time Taken(in seconds)
ANN	48%	494
CNN without batch normalisation and adaptive learning rates	70.8%	1397
CNN with adaptive learning rates and momentum	72.73%	1796
CNN with batch normalisation	76.67%	2000

Obeservations:

1) It can be seen from the above table least time is taken by ANN. The reason is the less nuber of epochs in the case of ANN. Number of epochs in case of ANN were 5 whereas the other had 20 epochs.

2) ANN has a very poor performance compared to CNN.

3) We can see from the above table with batch normalization the score is highest. Batch normalisation is a technique to that mitigates the effect of unstable gradients. The operation standardise and normalise the data of previous layer. Dropouts are regularisation technique that is used to prevent overfitting in the model.

4) If 5 convolutional layers and 3 FC layers are used in the CNN architecture the accuracy of the model can be further increased.

CNN model with batch normalisation and appropriate optimisers is the best architecture to train a model.

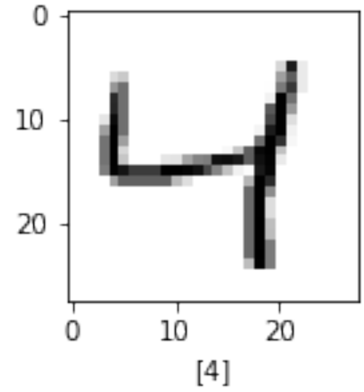
Task2

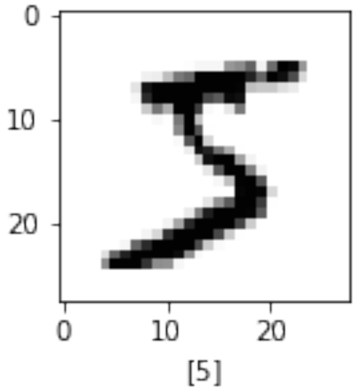
Task2 of the assignment is to train a model using CNN model on a dataset of our own choice. I have trained the CNN model on MNIST(“Modified National Institute of Standards and Technology”) of handwritten images.

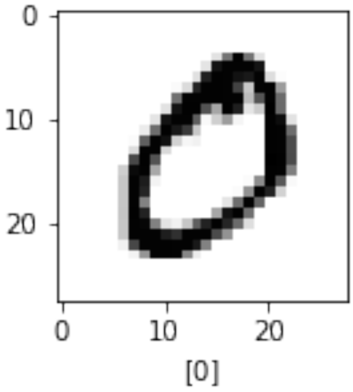
The dataset contains 60000 train and 10000 test greyscale images of size 28*28 pixel containing digits from 0-9.

A CNN model with 1 convolutional layer and 2 FC layer is used to train the model. With ‘Adam’ optimiser and 10 epochs the model gave an accuracy of 98.5% with very less loss.

Train Images:


0
10
20
0
10
20
[4]


0
10
20
0
10
20
[5]


0
10
20
0
10
20
[0]