HappyMonk report

Arvind Ganesh

June 2021

1 Introduction

Ada-Act is an activation function of the form $k_0 + k_1 x$ where k_0 and k_1 are adjustable parameters. We implement a neural network with 1 hidden layer on MNIST dataset using Ada-Act activation. The MNIST is a dataset of handwritten digits i.e. characters from 0-9. The network attempts to identify the value of digit from image. Our goal is to find the optimal value of k_0 and k_1 and analyze their behaviour.

The code base can be found in this link.

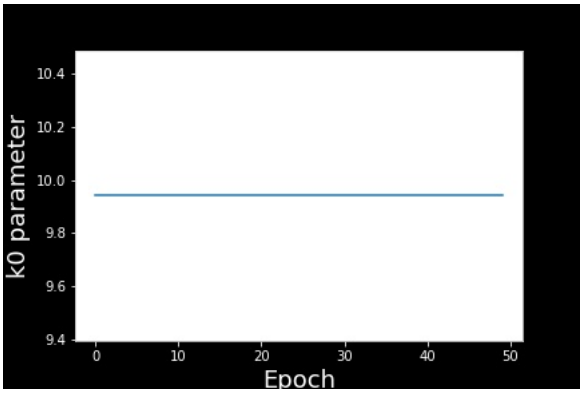
2 Details

We implement the network using Torch. The Ada-Act activation is implemented using the a custom class from autograd module. The code for the activation is as follows:

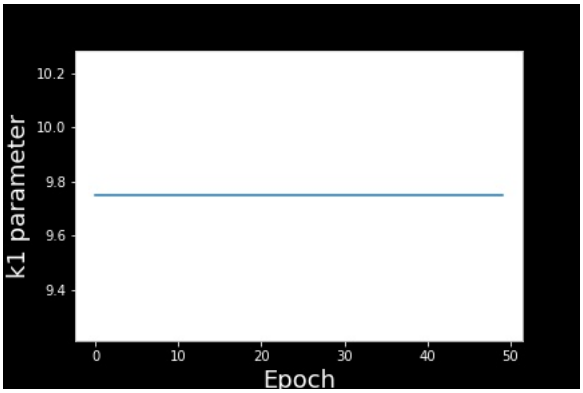
```
class learned_relu(nn.Module):
   def __init__(self, k0 = 10, k1 = 10):
      super().__init__()
      ones = torch.tensor([1.0])
      self.k0 = nn.Parameter(k0 * ones)
      self.k1 = nn.Parameter(k1 * ones)

   def forward(self, x):
      return self.k1 * x + self.k0
```

The 'nn.Parameter' ensures that k_0 and k_1 are defined in such a way that they are within the purview of the 'nn.Module' so that they may be considered as differentiable variables during backpropagation.

The network architecture is as follows:



(a) k0



(b) k1

Figure 1: Ada-Act parameters with epoch

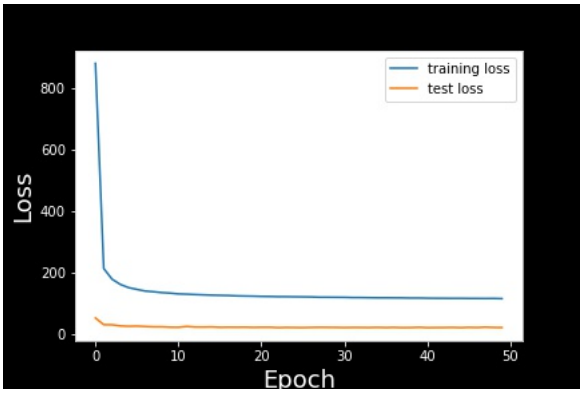
```
----------------------------------------------------------------
Layer (type)               Output Shape       Param #
================================================================
Linear-1                   [-1, 16]         12,560
learned_relu-2             [-1, 16]             0
Linear-3                   [-1, 10]           170
learned_relu-4             [-1, 10]             0
================================================================
Total params: 12,730
Trainable params: 12,730
Non-trainable params: 0
```

The two parameters of activation k_0 and k_1 are not considered in the above summary because module torchsummary merely lays out network parameters.

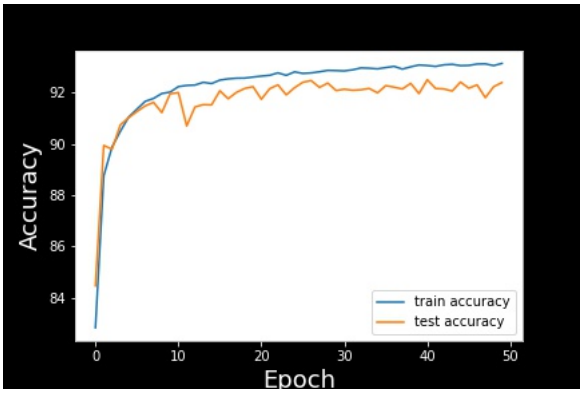
The initial values of k_0 and k_1 were manually tuned to be 10 and 10 respectively.

2.1 Results

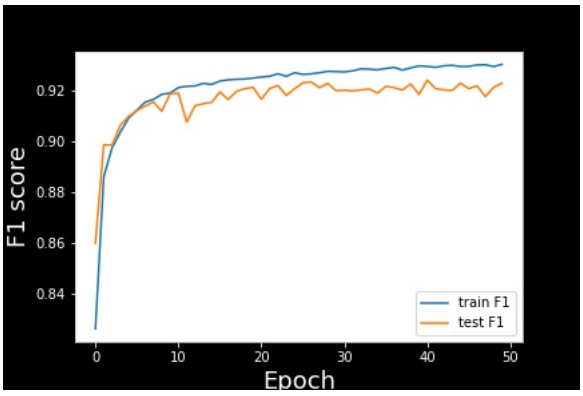
We plot the loss, accuracy and F1 score for training and test. The number of epochs was 50. Final values of activation parameters were $k_0 = 9.9567$ and $k_1 = 9.8420$ *which are obtained after backprop of single batch* as seen in Figure 1. **The final value of these two parameters is significantly different for different starting values**. But it does not affect the evaluted metrics seen in Figure 2.



(a) Loss



(b) Accuracy



(c) F1 score

Figure 2: Metrics evaluated at each epoch