CS898BD Deep Learning

Assignment2

10/17/2025

Submitted by,

Hasan Ahmed Shaik (N243V222)

Question 2 Report

1. Introduction and Approach

We evaluate and compare the performance of two activation functions **ReLU** (Rectified Linear Unit) and **Tanh** (Hyperbolic Tangent) with a 4 layer Convolutional Neural Network (CNN) architecture of image classification using CIFAR-10 dataset.
We measured how quickly each activation function allows the model to reach a specific evaluation threshold (training error of 25% or less)

Dataset and Processing :

- CIFAR-10 dataset consists of 60,000 (32 x 32) color images across 10 classes
- We split the original training dataset into
    - Training dataset : 40,000 images
    - Validation dataset : 10,000 images
- Normalize the pixel values from [0,255] to [0.0,1.0]
- Convert labels using one-hot encoding for categorical cross-entropy loss

Model Architecture :

The model architecture consists of 4 layered CNN for both activation functions used and they are as shown below :

1st layer : Conv2D(32 filters) using Batch Normalization with Maxpooling and Dropout=0.2
2nd layer : Conv2D(64 filters) using Batch Normalization with Maxpooling and Dropout=0.2
3rd layer : Conv2D(128 filters) using Batch Normalization with Maxpooling and Dropout=0.2
Flatten layer : Dense layer with 128 units, Batch Normalization with Dropout =0.3 with output of10 units with activation function = SoftMax

Evaluation Metrics and Stopping Criterion :

Training Error Rate = 1 – Training Accuracy(per epoch)

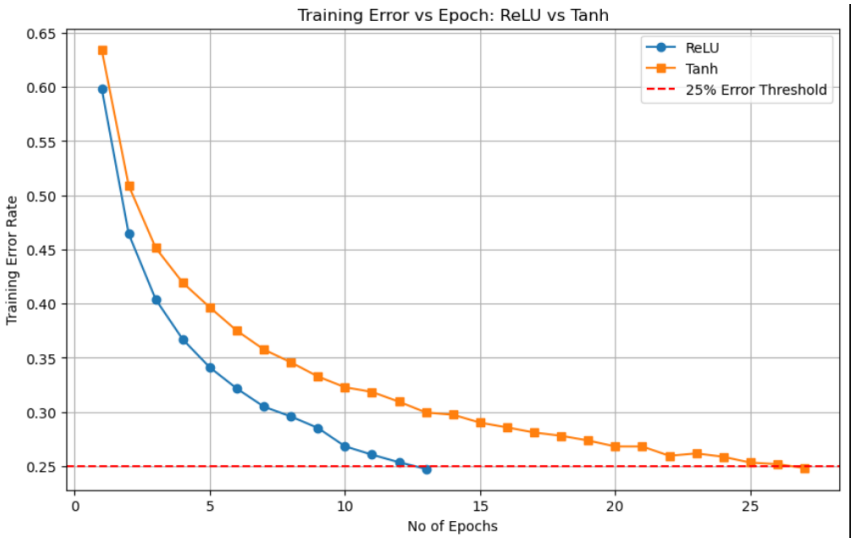
Training stopped once the Training error dropped <=25%

For each activation function we have recorded the following :

- Number of epochs required to reach the training errors threshold
- Training error per epoch
- Time taken per epoch (in seconds)

Results :

1. Graph : Training Error vs Epoch (ReLU vs Tanh)



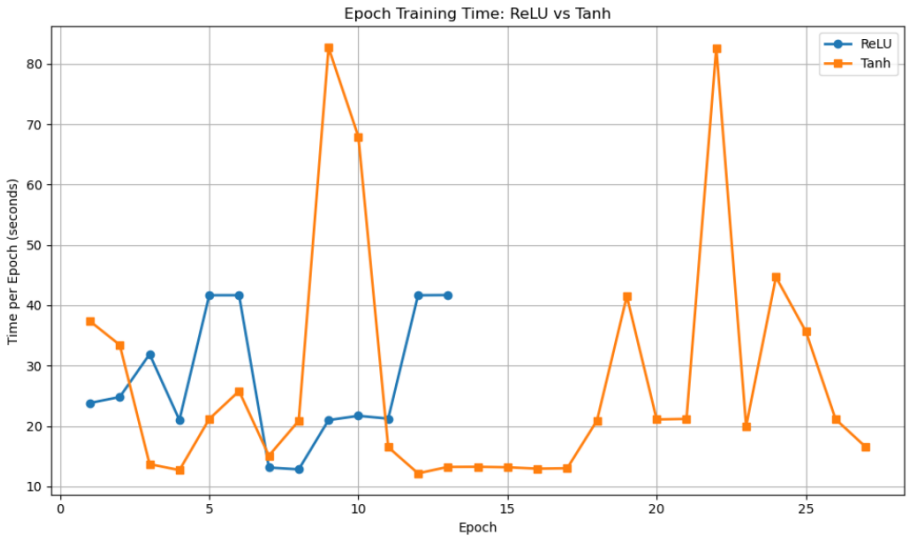
ReLU :

Started with training error of 60 % and rapidly decreasing crossing the threshold at Epoch 13. At Epoch 13 reaching training error of ~24.7%

Tanh :

Started with higher training error than ReLU (~64%) and it converged slowing than ReLU reaching 25% threshold at Epoch 26. At Epoch 26 training error of ~24.9%

2. Time taken to complete per Epoch



ReLU :

The average time per epoch is between 20-40 seconds(approx). Most epochs completed under 30 seconds.

Tanh :

Several epochs took over 60-80 seconds. Many epochs showed spikes in duration and average time per epoch is (~ 30- 50) seconds.

Observation :

- ReLU's simple **(max(0,x))** operation avoids vanishing gradient problem that Tanh problem has in deep networks. Tanh saturates at both ends causing gradients to become extremely small which shows it learning slow compared to ReLU.
- ReLU only has threshold operation making it computationally efficient than Tanh which requires exponential calculation.
- In deeper layers, ReLU allows gradients to flow more freely during backpropagation. Tanh's bounded output range can lead to diminishing gradients, especially in early epochs as seen as it's slower error reduction curve in graph

- We can see that there are more spikes in Tanh's training time in epoch which says that more memory usage and computational resources are used when compared to ReLU.

Conclusion

ReLU has faster convergence (13 epochs) when compared to Tanh (26 epochs). ReLU is more efficient in computational resources used and it provides more stable training times than Tanh, which has few spikes in epochs