ASSIGNMENT – I

DEEP LEARNING

Roll Number:160122737152

Overview:

We trained a feedforward neural network on the Fashion MNIST dataset, experimenting with different hyperparameter combinations to determine best performing model. The evaluation was based on validation and test accuracy.

Observations and Interpretations:

1. Impact of hidden layers and neurons:

Model with moderate depth i.e **3-4 layers** performed better than more deep networks. Too many layers (5 layers) slightly degraded the performance, possibly due to overfitting.

Using **64 or 128 neurons** per layer produced better results than using 32 neurons per layer.

1. Effect of Optimizers:

**Adam and RMSprop** performed better in terms of accuracy. Adam is one of the best optimizers in deep learning, combining momentum and adaptive learning rates.

Momentum-based methods (Nesterov, Momentum SGD) improved SGD’s performance, but were still slightly behind Adam.

1. Learning rate & Weight Decay:

A learning rate of **0.001** worked better whereas 0.0001 resulted in slower training.

L2 regularization (**0.0005**) improved generalization, but a high value (0.5) restricted learning.

1. Activation Function:

**ReLU** outperformed Sigmoid due to its ability to prevent vanishing gradients.

Sigmoid networks showed slower convergence and slightly lower accuracy.

1. Weight Initialization:

**Xavier** (Glorot) initialization helped achieve stable training and better accuracy compared to random initialization.

Recommended Hyperparameter Configurations for MNIST

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Hidden layers | Optimizer | Learning rate | Activation | Weight decay | Weight init | Batch size |
| 1 | [64,64,64] | Adam | 0.001 | ReLu | 0.0005 | Xavier | 32 |
| 2 | [128,128,128] | RMSprop | 0.001 | ReLu | 0 | Xavier | 64 |
| 3 | [64,64,64,64] | Nesterov | 0.001 | ReLu | 0.0005 | Xavier | 32 |

These configurations perform well on MNIST. Applying these learnings to MNIST, we can expect test accuracies up to 97%, as MNIST is considered simpler than Fashion MNIST.