

SML Assignment - 1

Ans 2

Hidden layers used: 3

Hyperparameters used:

Learning rate = 0.01

Momentum = 0.9

Output dimension : 10

No. of epochs : 30

Batch-size : 512

Ans 3

Loss function used: MSE (Mean-Squared Error)

We use this loss function since minimizing the MSE is equivalent to applying Maximum likelihood to our model, giving us better results.

Also, the data used is continuous and not discrete, since BCE (Cross-Entropy loss) works better on discrete data but MSE works better on such continuous data.

Here, by continuous data, we mean images. Had it been on labels, which are discrete we would've gone for multiclass entropy loss.