CS 725: Foundations of Machine Learning Homework 2

Soumen Kumar Mondal
23m2157@iitb.ac.in

Naay Balodia
23m2166@iitb.ac.in

September 3, 2023

Abstract

In this assignment, we will implement Feed Forward Neural Network model on a toy datasets for 4 class simple classification and 10 class digits classification. Our key goal in this assignment is to correctly implement these models and analyze the results we obtained.

1 Classification on Simple Dataset

1.1 Learning Rate

The learning rate is a crucial hyper-parameter in machine learning algorithms, especially in gradient-based optimization methods like gradient descent. It determines the step size at which the algorithm updates the model's parameters during training. A higher learning rate can lead to faster convergence, but it might also cause overshooting and divergence. Conversely, a lower learning rate can lead to slower convergence, potentially getting stuck in local minima. Finding the right balance is essential for achieving optimal training performance and avoiding convergence issues. Experimenting with different learning rates helps identify the optimal rate that maximizes the model's accuracy and minimizes its loss during training and validation.

The impact of varying learning rates on accuracy and loss was explored for the simple classification model. The outcomes of these experiments can be summarized as follows:

Accuracy: Analyzing Figure 1 reveals that the FFNN model achieves its highest accuracy when the learning rate is set to 0.125. Since the experiment is conducted with two choice of epoch, it can be seen that the trend of learning rate is similar for both the choice of number of epoch. Therefore, we will select the learning rate as 0.125 as it will require less number of epochs for convergence.

Loss: From Figure 2, it is evident that the loss is minimum and achieves convergence when the learning rate is set to 0.125. The trend of loss is also similar in both choice of epochs. Hence, the learning rate of 0.125 can be adopted for simple classification model.

In summary, the learning rate of 0.125 is optimal for this simple classification model as it reaches to convergence in less number of epochs without having significant impact on the accuracy and loss of the FFNN model. **Therefore, we select learning rate of 0.125 as the best learning rate for the simple classification FFNN model.**

1.2 Number of Epoch

The number of epochs in machine learning signifies the count of times the entire data set is iterated through during training. The choice of the number of epochs plays a pivotal role in model performance. Too few epochs may lead to under-fitting, where the model hasn't learned enough from the data, while too many epochs might cause over-fitting, where the model captures noise in the training data, failing to generalize well to new data. Striking the right balance by monitoring validation performance helps achieve optimal model training and prevent over-fitting or premature convergence.

The number of epoch are varied while keeping the learning rate same as 0.125 which is the best value of learning rate. It can be observed on Figure 3 and Figure 4 that the number of epoch 100 is optimal as it reaches convergence early. Nevertheless, the difference in accuracy and loss in both training and validation data set between epoch 100 and epoch 250 is insignificant. **Therefore, we select number of epoch of 100 as the best number of epoch for the simple classification FFNN model.**

The FFNN hyper-parameters that are selected for the simple classification model is shown in Table 3. Moreover, the FFNN architecture for the simple classification is shown in Table 1.

Layers:	input layer ⟶ hidden layer 1 ⟶ hidden layer 2 ⟶ output layer
Neurons (or Units):	2 features ⟶ 32 neurons ⟶ 16 neurons ⟶ 4 classes
Activation Function:	ReLU in all layers
Optimizer:	Stochastic Gradient Descent with momentum of 0.9

Table 1: FFNN hyper-parameters selected for the simple classification model

2 Classification on Digits Dataset

2.1 Learning Rate

The importance of learning rate in the ML model is discussed in Section 1.1. The impact of varying learning rates on accuracy and loss was explored for the digits classification model. The outcomes of these experiments can be summarized as follows:

Accuracy: Analyzing Figure 5 reveals that the Linear Classifier (LC) model achieves its highest accuracy when the learning rate is set to 0.025. The number of epoch has insignificant effect on learning rate. As expected, when the learning rate is too small, the accuracy will drop at lower number of epochs which is evident in the same plot.

Loss: From Figure 6, it is evident that the loss is minimum and achieves convergence in lowest number of epochs when the learning rate is set to 0.025. As expected, a lower learning rate will require more number of epochs to achieve minimum loss. The difference between the losses in the training dataset is insignificant.

In summary, the learning rate of 0.025 is optimal for this digits classification FFNN model as it reaches to convergence in less number of epochs without having significant impact on the accuracy and loss of the model. **Therefore, we select learning rate of 0.025 as the best learning rate for the digits classification model.**

2.2 Number of Epoch

The importance of number of epoch in the ML model is discussed in Section 1.2. The impact of varying number of epoch on accuracy and loss was explored for the digits classification model. The number of epoch are varied while keeping the learning rate same as 0.125 and 0.025. It can be observed on Figure 7 and Figure 8 that the number of epoch 250 is optimal as it reaches convergence early and it has slightly better accuracy than the number of epoch 100. Nevertheless, the difference in accuracy and loss in both training and validation data set between epoch 250 and epoch 100 is insignificant. However, we choose number of epoch 250 as best because of the uncertainty in unseen data where number of epoch 100 might not be sufficient to reach the convergence. **Therefore, we select number of epoch of 250 as the best number of epoch for digits classification FFNN model.**

The best hyper-parameters that are selected for the digits classification FFNN model is shown in Table 3. Moreover, the FFNN architecture for the digits classification is shown in Table 2.

Layers:	input ⟶ hidden 1 ⟶ hidden 2 ⟶ hidden 3 ⟶ output
Neurons (or Units):	64 features ⟶ 64 unit ⟶ 32 unit ⟶ 16 unit ⟶ 10 classes
Activation Function:	ReLU in all layers
Optimizer:	Stochastic Gradient Descent with momentum of 0.9

Table 2: FFNN hyper-parameters selected for the digits classification model

3 General Points

3.1 Best Epoch and Learning Rate

The FFNN hyper-parameters that are selected for the simple classification model and digits classification model is shown in Table 3.

Model Name	Best Learning Rate	Best Number of Epoch
Simple Classification	0.125	100
Digits Classification	0.025	250

Table 3: Best hyper-parameters selected for the simple and digits classification model

4 Data Pre-processioning

For both simple and digits dataset, no preprocessing is performed. **Hence it is ”Not Done”.**

5 Prevention of Overfitting

Preventing overfitting in a FFNN is a crucial aspect of building robust machine learning models. One effective technique for addressing overfitting in FFNNs is L2 regularization, also known as weight decay. L2 regularization adds a penalty term to the loss function, discouraging the model from assigning excessively large weights to specific features during training. By doing so, L2 regularization encourages the model to use all features more evenly and reduces the risk of fitting noise or outliers in the training data. This regularization technique helps in achieving a better balance between model complexity and generalization performance, leading to improved model performance on unseen data. By tuning the strength of the L2 regularization term, we can strike the right balance between fitting the training data well and avoiding overfitting, resulting in more reliable and robust FFNN models for various machine learning tasks.[1, 2]

The L2 regularization equation is given in the following equation below:

$$\min_{w,b} \left[L(w,b) + \frac{\lambda}{2} \cdot \|\mathbf{w}\|^2 \right] \text{ for } \lambda > 0 \tag{1}$$

For our FFNN model, we have tuned the hyperparameter λ as 0.0015.

References

[1] Andrew Ng. *Deep Learning Specialization*. 2017. URL: https://youtube.com/playlist?list=PLpFsSf5Dm-pd5d3rjNtIXUHT-v7bdaEIe&si=pYyzJQk7P5FFmqGK.

[2] Aston Zhang et al. *Dive into Deep Learning*. 2021. URL: https://d2l.ai/.

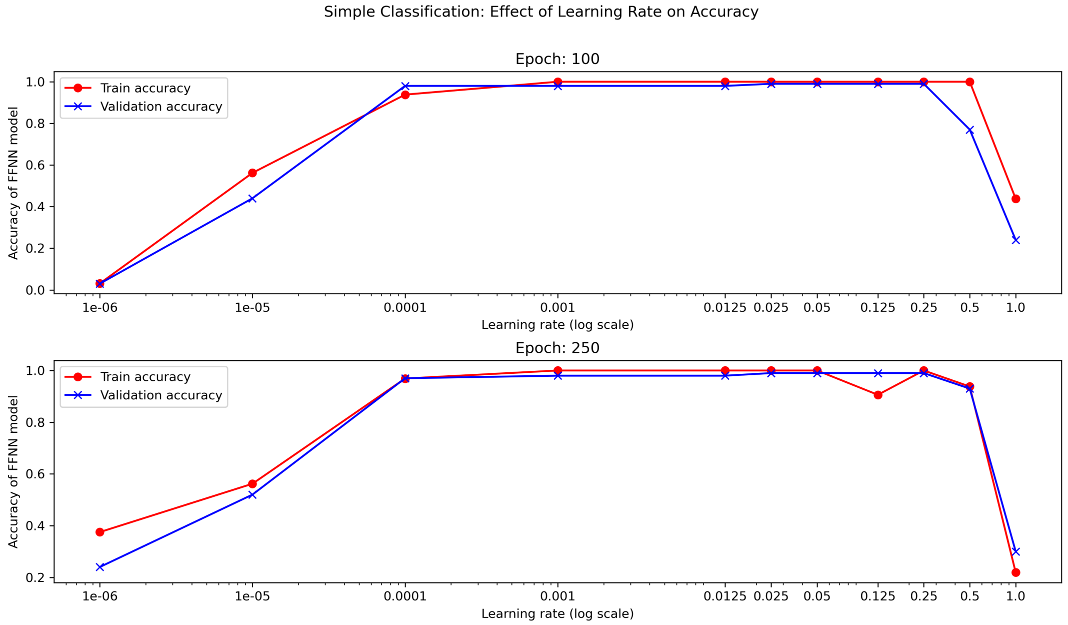

Figure 1: Effect of learning rate on accuracy of simple classification model

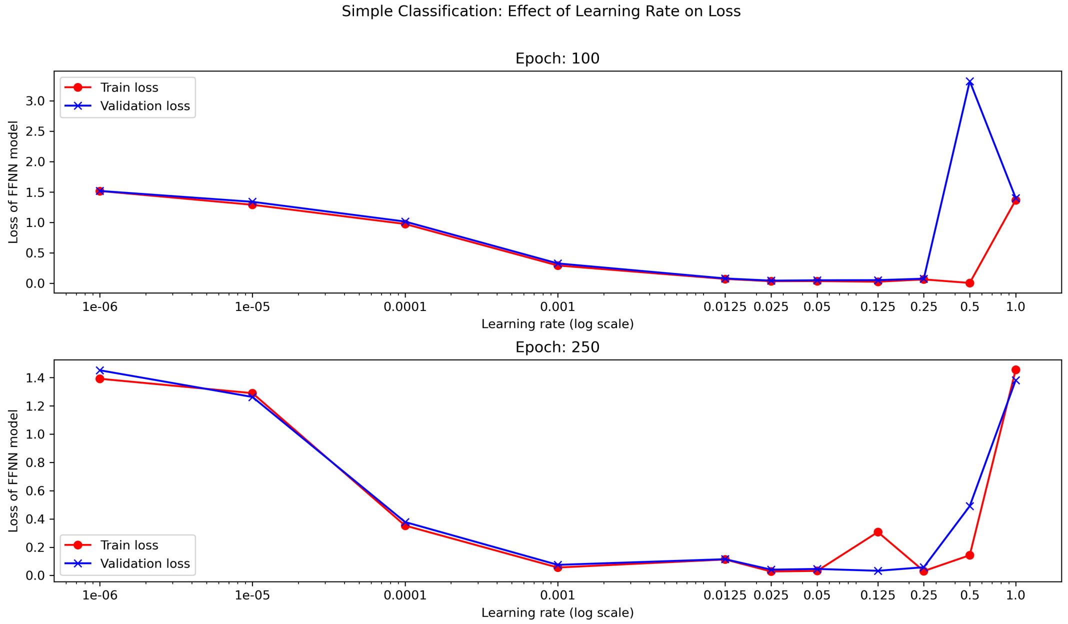

Figure 2: Effect of learning rate on loss of simple classification model

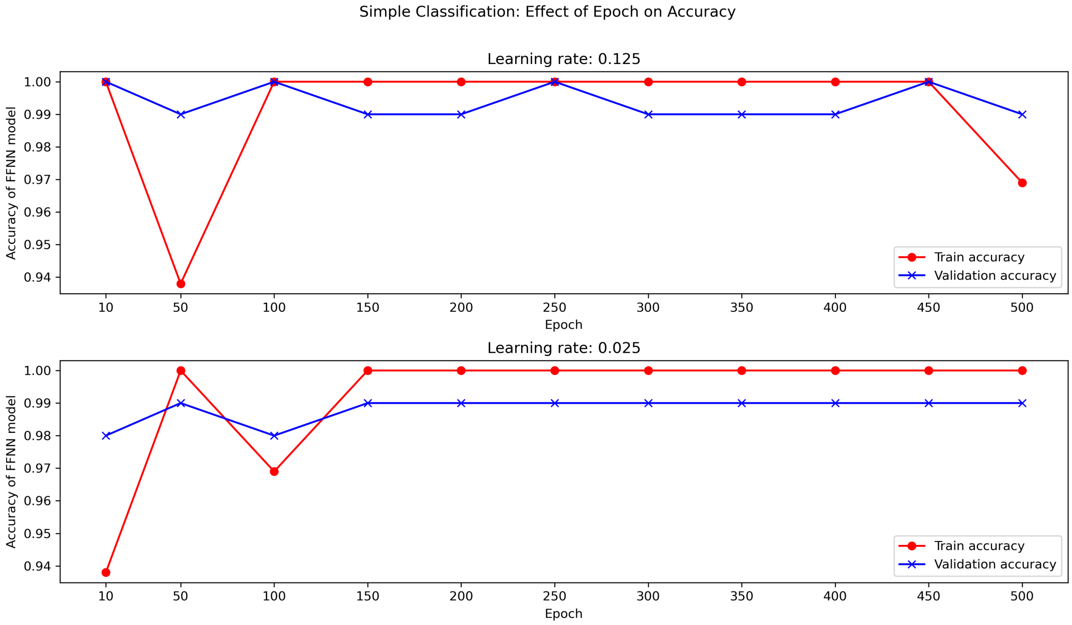

Figure 3: Effect of epoch on accuracy of simple classification model

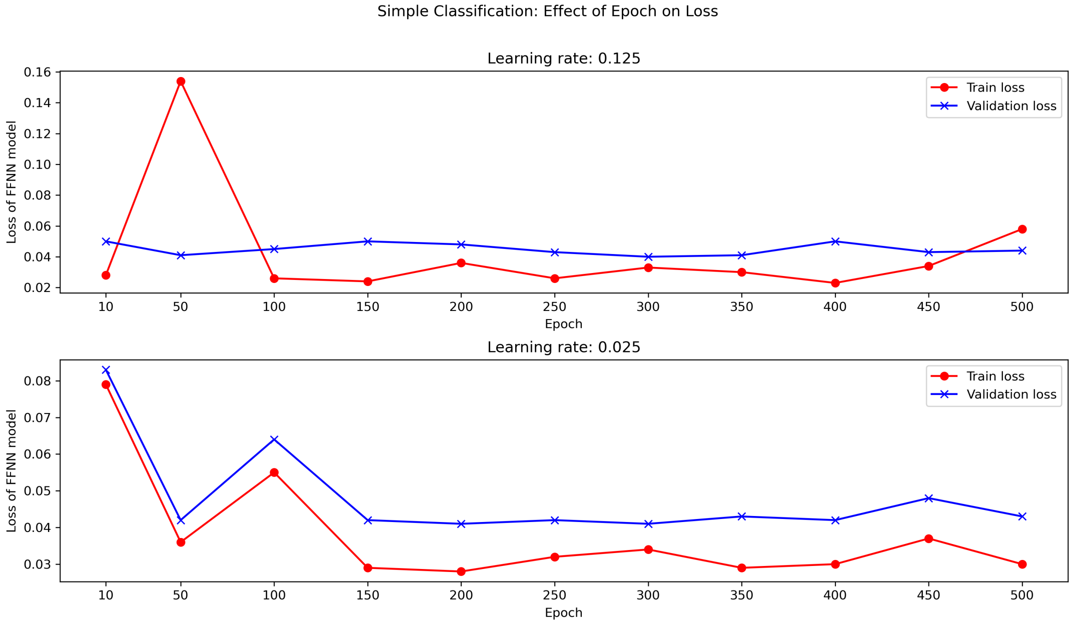

Figure 4: Effect of epoch on loss of simple classification model

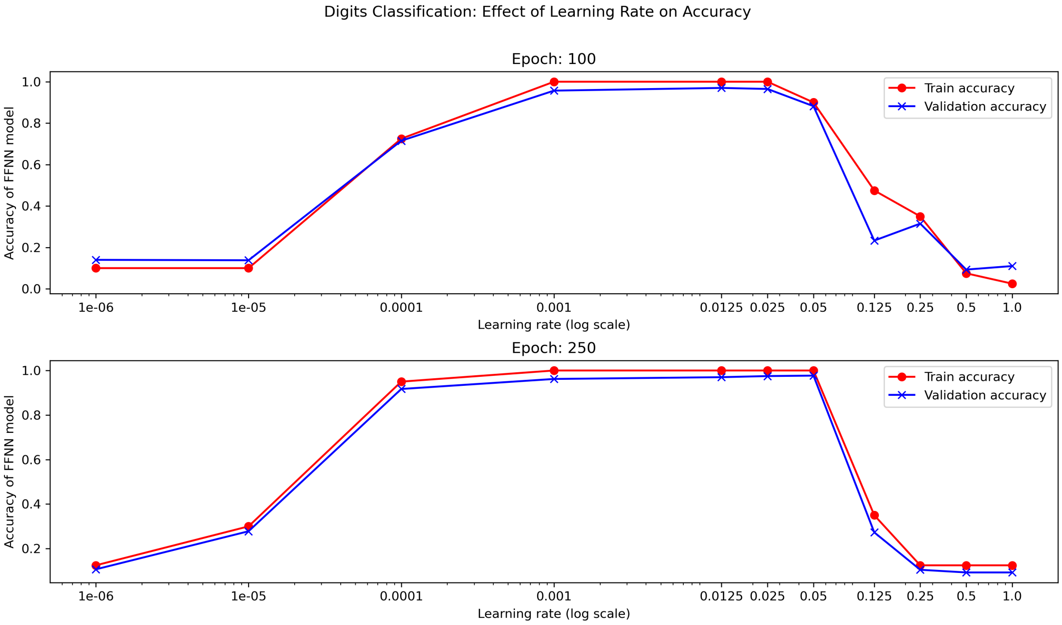

Figure 5: Effect of learning rate on accuracy of digits classification model

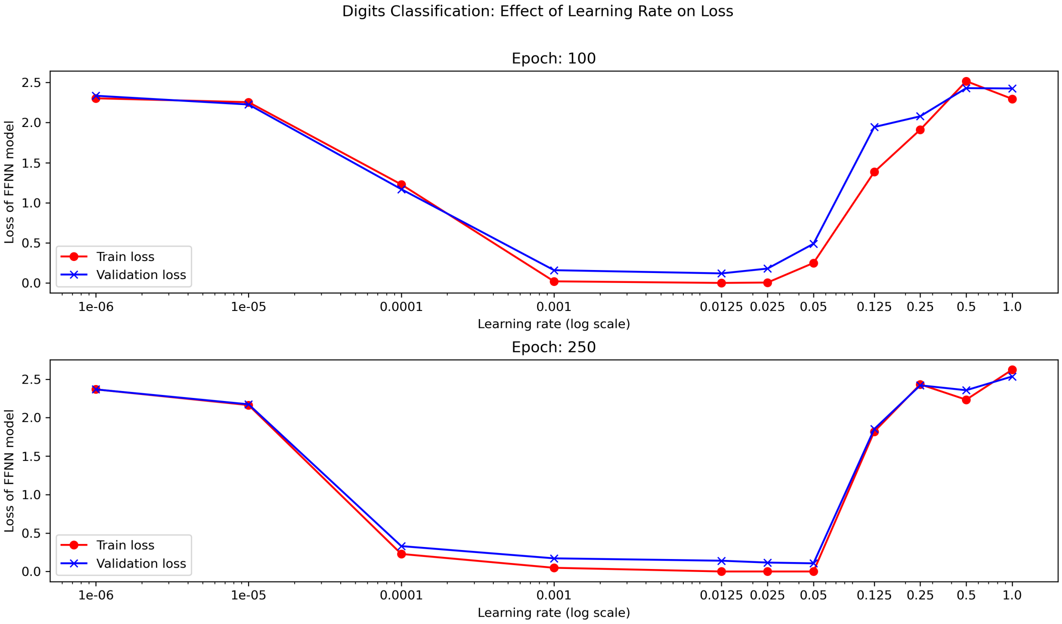

Figure 6: Effect of learning rate on loss of digits classification model

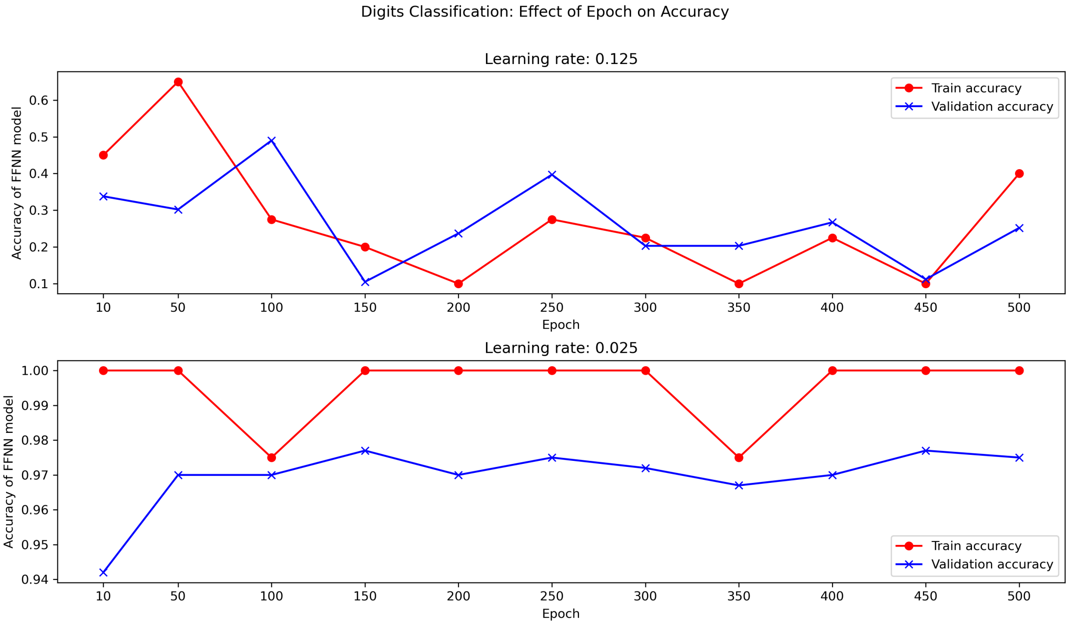

Figure 7: Effect of epoch on accuracy of digits classification model

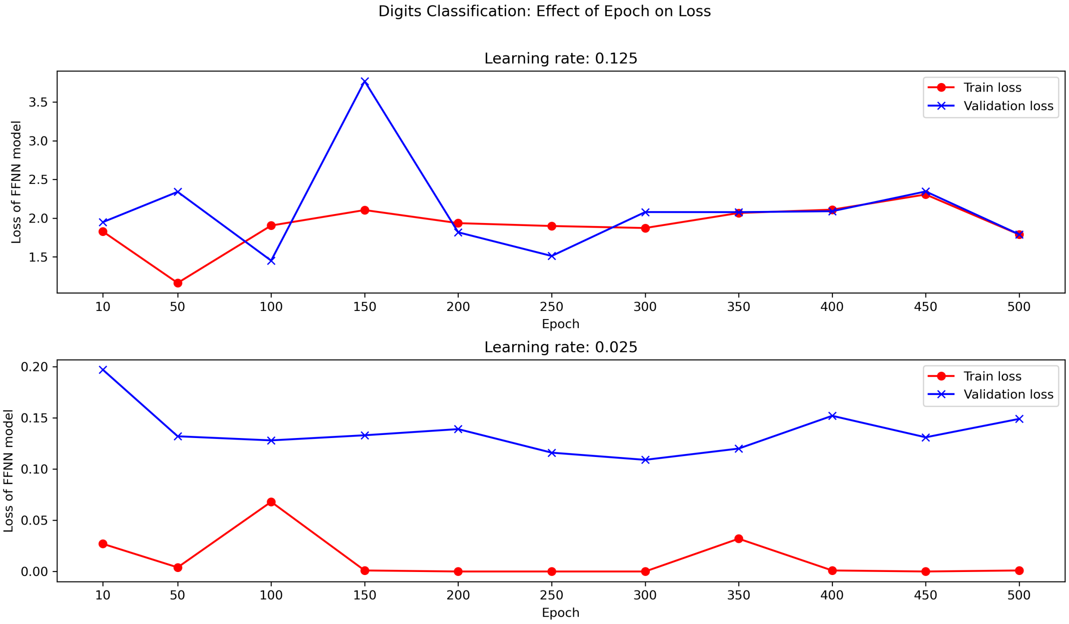

Figure 8: Effect of epoch on loss of digits classification model