CENG499

Homework 1

Part 3

Berke Sina Ahlatcı

2468502

Did a grid search over hyperparameter values 10 times and used averages of accuracy and loss values in order to find the best hyperparameter configuration.

1 Hyperparameters

I performed an extensive grid search over the values :

- epochs = [40, 80]
- learning rates = [0.1, 0.001]
- layer numbers = [2, 3]
- layer sizes = [8, 32]
- activation functions = [nn.ReLU(), nn.Sigmoid()]

2 Accuracy Results on The Validation Dataset

Accuracy table with hyperparameter configurations and corresponding (accuracy percentage $\pm$ confidence interval radius) values on the validation set.

Layers/Epochs/LayerSizes/Activator	0.1	0.001
2 / 40 / No Hidden Layer	51.70% +- 0.98%	82.42% +- 0.75%
2 / 80 / No Hidden Layer	59.80% +- 0.96%	84.11% +- 0.72%
3 / 40 / 8 / ReLU()	47.13% +- 0.98%	67.05% +- 0.92%
3 / 40 / 8 / Sigmoid()	46.97% +- 0.98%	48.58% +- 0.98%
3 / 40 / 32 / ReLU()	46.14% +- 0.98%	76.42% +- 0.83%
3 / 40 / 32 / Sigmoid()	60.64% +- 0.96%	64.96% +- 0.94%
3 / 80 / 8 / ReLU()	43.69% +- 0.97%	62.81% +- 0.95%
3 / 80 / 8 / Sigmoid()	39.57% +- 0.96%	58.76% +- 0.97%
3 / 80 / 32 / ReLU()	56.76% +- 0.97%	79.50% +- 0.79%
3 / 80 / 32 / Sigmoid()	59.15% +- 0.96%	70.74% +- 0.89%

Figure 1: Accuracy values on the validation set.

3 Additional Q/A

- How could one understand that a model being trained starts to overfit? When training loss becomes so small and validation loss starts to grow, we can assume that modes starts to memorize the data.
- Is there a "best" learning rate? Best accuracy acquired with the learning rate of 0.001. Additionally, mean of accuracy values obtained with 0.001 learning rate is bigger than 0.1 (0.1's mean is 0.51, 0.001's mean is 0.69).
- Is there a "best" activation function? Best accuracy acquired with the activation function of ReLU(). Additionally, mean of accuracy values obtained with ReLU activation function is bigger than 0.1 (Sigmoid's mean is 0.56, ReLU's mean is 0.60).

4 Result

Best hyperparameter configuration is:

- epoch number = 80
- learning rate = 0.001
- layer number = 2 , hence there isn't any layer size and activation function (no hidden layers).

After the final training with this hyperparameter configuration, loss and accuracy values are as follows:

- Loss value on validation data is 1.83 and accuracy value on validation data is 76%.
- Loss value on test data is 1.83. Accuracy value on test data is 78%.

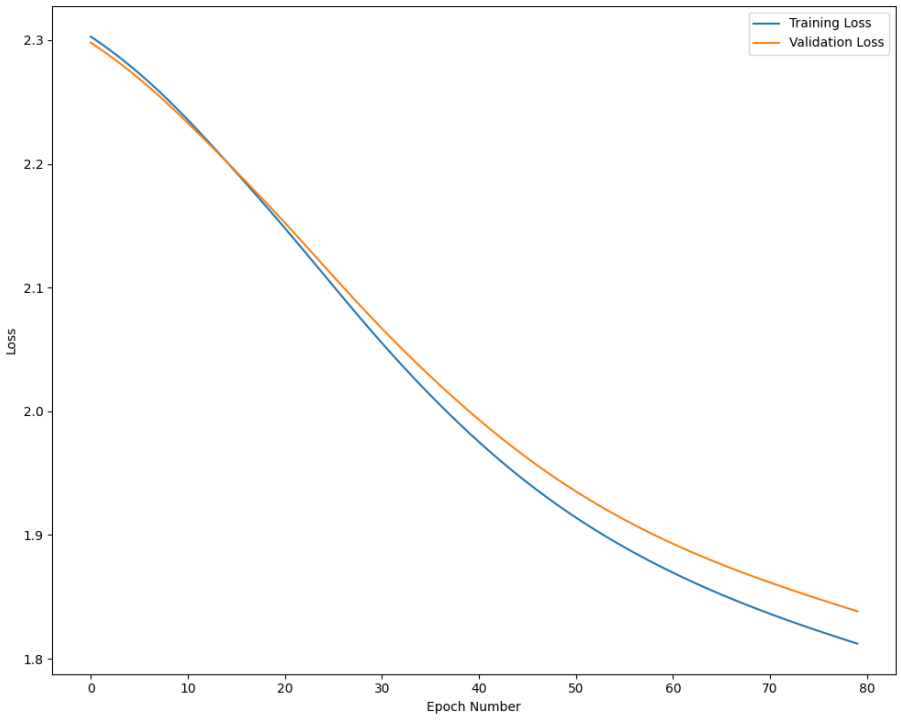

Figure 2: Loss plot on validation and training datasets.