Ceng499 HW1 Report - Part 3

Atınç Utku Alparslan
2380061

1 Part 3

In this part, I implemented an MLP training code for classification task. For tuning the hyper parameters, I implemented an extensive grid search. The parameters are tested are:

- Activation Functions: Sigmoid and Tanh
- Layer Numbers: 2 and 4
- Layer Sizes: 8 and 132
- Learning Rates: 0.1 and 0.0001
- Epoch Numbers: 30 and 80

I wanted to test more parameters. However, due to the constraints of my computer, I could not be able to extend it further. I tried to test relatively small and big numbers in order to show the difference more clearly. The accuracy table with the tested hyperparameters below:

Layer No:	Epoch No:	Layer Size:	Activation Func:	When LR: 0.1	When LR: 0.0001
2	30	No Hidden Layer	No Hidden Layer	67.06% +- 0.92%	53.07% +- 0.98%
2	80	No Hidden Layer	No Hidden Layer	60.34% +- 0.96%	68.72% +- 0.91%
4	30	8	Sigmoid	47.72% +- 0.98%	10.55% +- 0.60%
4	30	8	Tanh	59.88% +- 0.96%	23.09% +- 0.83%
4	30	132	Sigmoid	28.47% +- 0.88%	21.65% +- 0.81%
4	30	132	Tanh	55.02% +- 0.98%	65.53% +- 0.93%
4	80	8	Sigmoid	51.50% +- 0.98%	15.79% +- 0.71%
4	80	8	Tanh	46.49% +- 0.98%	35.65% +- 0.94%
4	80	132	Sigmoid	24.07% +- 0.84%	40.69% +- 0.96%
4	80	132	Tanh	71.23% +- 0.89%	77.75% +- 0.82%

Best loss value is 1.75 and the best hyper parameters among tested hyper parameter configurations are:

- Activation Functions: Tanh
- Layer Number: 4
- Learning Rate: 0.1
- Epoch Number: 80
- Layer Size: 132

Notes:

- The model being trained starts to overfit when accuracy is high, but the loss value is also high. That means the model starts to memorize the data. Therefore after testing, it fails to label values correctly.
- The large learning rate is useful as it allows to model learn faster. However, it has a cost of arriving on a sub-optimal final set of weights.
- The small learning rate allows model to achieve more optimal results. However, it affects the training time, making it longer than needed.
- Among the chosen test parameters, the best loss value and accuracy is gathered with the Tanh function and 0.1 learning rate among all tested values.

- Validation loss is 1.85 and validation accuracy is 0.61.
- Train Loss is 1.78 and train accuracy is 0.68.
- Test loss is 1.83 and test accuracy is 0.62

The plot of validation loss and training loss is below: