DD2424 Assignment 2 Report (For Optional Part)

Zesen Wang

1 Optimize the Performance of the Network

In this part, I implement 3 optimization method.

(a) Early stopping: Save the model with best validation accuracy, and stop training if there is no improvement on validation set for more than 10 epoches.

(b) He Initialization: Set the standard deviation as $\sqrt{2/n_{in}}$

(c) Do a more exhaustic search to find good value with He Initialization.

(d) Decay the learning rate by 0.1 after every 8 epoches.

For the following experiments, these parameters are same.

```
lambda=0.0023292248102687557, eta=0.017453577972249945, momemtum=0.95, n_batch=100
```

```
Optimization method: He Initialization + Exhaustic Random Search
n_epoch=30
Accuracy on test set: 0.5151
```

```
Optimization method: Early Stopping + Exhaustic Random Search
N_no_improvement=10
Accuracy on test set: 0.5179
```

```
Optimization method: Learning Rate Decay by 0.1 + Exhaustic Random Search
n_epoch=30, Decay_cycle=8
Accuracy on test set: 0.5200
```

```
Optimization method: All Methods
Decay_cycle=8, N_no_improvement=10
Accuracy on test set: 0.5241
```

```
Optimization method: None
Accuracy on test set: 0.5127
```

It indicates that the Learning Rate Decay improves the network most, and their combination gives the most gain among all experiments.

2 Leaky ReLU Activation

The activation function I use here is leaky ReLU activation function.

$$\varphi(x) = \begin{cases} x(x \geq 0) \\ 0.01x(x \leq 0) \end{cases}$$

I tests the leaky ReLU function on the experiment with no optimization method and the experiment with all optimization method mentioned above. The results are listed below.

```
Optimization method: All Methods + Leaky ReLU Activation
Decay_cycle=8, N_no_improvement=10
Accuracy on test set: 0.5246

Optimization method: None + Leaky ReLU Activation
Accuracy on test set: 0.5210
```

It shows that leaky ReLU activation actually leads to some improvement for the accuracy on test set compared to simple ReLU activation.