CS547-HW1

Name: Qiyang Chen
ID: qiyangc2

September 2019

1 Introduction

In HW1 we are asked to train a neural network for MNIST dataset. I have used 2 methods to complete this homework.

1.1 Plan A

In Plan A: I trained a MLP model which has one hidden layer made up of 200 hidden units. And the Test Accuracy of this model is 97.37%. I set all weights to zero at first and use SGD to optimize the weights. The learning rate is 0.1 at first. I trained 50 epochs to achieve 97.37%.

1.2 Plan B

In Plan B: I trained three smaller models than what I did in Plan A. They have 64,60 and 100 hidden units respectively. Of course, theses three models only have one hidden layer. And other hyper-parameters is the same as Plan A but in this Plan I ensemble the these three models' output and feed it into argmax function as the final prediction. And I also use the momentum as well which is set as 0.9. Plan B can achieve 98.23% in final.

2 Conclusion

I prefer to train 3 small models and ensemble. Because we have less units in Plan B to train than Plan A. But, of course, in HW1, I don't think the training time of these two plans is a very big difference. Because, compared with more complicated models, both of them are "small". That's all what I did in my HW1.