CS547-HW3

Name: Qiyang Chen
ID: qiyangc2
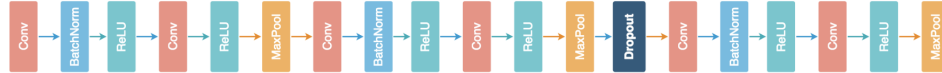
September 28, 2019

1 Introduction

In HW3 we are asked to train a deep convolution network on a GPU with PyTorch for the CIFAR10 dataset.

2 Architecture

My main architecture is shown as below:

I set 3 blocks in CNN model. In each block, I set 2 CNN layers, and I use batch normalize and ReLU between 2 CNN layers. And in each block, I also use max pooling layer to reduce the feature dimension which can help extract features better and compute. After CNN, I build MLP to project the image feature map from CNN blocks. The architecture of MLP is shown as below:

In MLP I also set dropout layer to avoid over fitting. After the MLP, the output is the probability distribution of 10 classes. And other parameters are shown in the table below:

learning rate = 0.001
weight decay = 5e-4
epochs = 50
$batch_size = 64$

3 Result

I trained this model on AWS by using p2.xlarge. I can get 85.09% by using this architecture with Adam Optimizer. And I can get 85% by using this architecture

with RMSProp Optimizer.

4 Conclusion

We can figure out RMSProp and Adam have the similar performance. But in fact, I think I should try SGD and layer-dropout which can improve over-fit on neural networks like L2 Regularization. That's all what I did in my HW3.