CSE 455 Homework 5

Yoshihiro Kumazawa

August 22, 2020

1 Installing PyTorch

Done!

2 Find the best network

2.1 Training a classifier using only one fully connected Layer

See Figure 1a. We can say that the the model successfully trained since the loss is decreasing throughout the training process and there is a healthy gap between the training accuracy and testing accuracy.

2.2 Training a classifier using multiple fully connected Layers

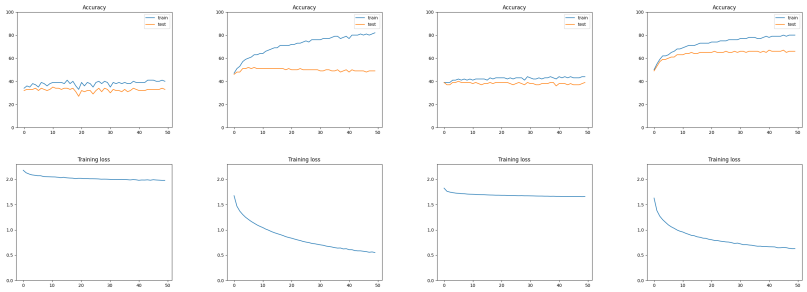
See Figure 1b. The training is not successful because the testing accuracy plateaus whereas the training keeps increasing.

2.2.1 Question

See Figure 1c. The model accuracy is significantly worse than the previous model. This is because the model is expressively limited since it has less non-linearity. The model can actually become just as good as LazyNet since without activations, the forward pass is just a couple of matrix multiplications, which is nothing but a single matrix multiplication by the composed matrices. But somehow, by separating the weight update process in back propagation, it achieves a slightly higher accuracy than LazyNet.

2.3 Training a classifier using convolutions

Our CoolNet is based on LeNet [2]. In addition, batch normalization [1] is performed after each convolutional layer and linear layer except the final layer in order to get the network to converge faster. As a result, the network got higher accuracy then the previous models. See Figure 1d for the training result.



(a) (b) (c) (d)

Figure 1: Training results of LazyNet, BoringNet and CoolNet: (a) LazyNet, (b) BoringNet, (c) BoringNet without activations, (d) CoolNet with batch size 4.

2.3.1 Question

3 How does learning rate work?

4 Data Augmentation

5 Change the loss function

References

[1] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

[2] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.