

人工神经网络·第二次作业报告

吴明恒

2018011288

日期：2020 年 10 月 17 日

1 背景

在实验给定的框架的基础上，本次作业需要实现 Dropout 与 Batch normalization，并将其应用到两种神经网络当中，它们分别是 MLP 与 CNN。

在完成 Dropout、Batch normalization 与神经网络的设计之后，我们需要在 CIFAR-10 数据集上对其进行评价，以及对不同的 Dropout rate 设置进行性能对比与分析，并画出 Loss 与 Accuracy 的变化图像。

2 原理

2.1 Dropout

Dropout 的原理是在前向传播的时候，让某个神经元的激活值以一定的概率 p 停止工作，这样可以使模型泛化性更强，因为它不会太依赖某些局部的特征。同时为了不变化输入的分布，我们需要再除以一个系数。

2.2 Batch normalization

Batch normalization 的原理是通过改变数据的分布状态（归一化），使得神经网络的每一层数据都可以较好地保持类 0-1 正态曲线的分布，它可以很好地解决 Internal Covariate Shift 问题，降低梯度消失的发生概率，使得模型的训练更好进行。

3 实验

3.1 模型设置

3.1.1 MLP

```
Model(  
  (linear): Sequential(  
    (0): Linear(in_features=3072, out_features=512, bias=True)  
    (1): BatchNorm1d()  
    (2): ReLU()
```

```

        (3): Dropout()
        (4): Linear(in_features=512, out_features=10, bias=True)
    )
    (loss): CrossEntropyLoss()
)

```

3.1.2 CNN

```

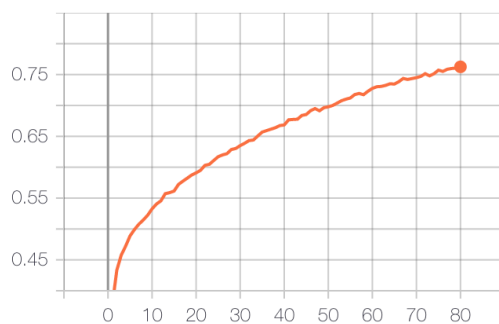
Model(
  (conv1): Sequential(
    (0): Conv2d(3, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d()
    (2): ReLU()
    (3): Dropout()
    (4): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (5): Conv2d(128, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (6): BatchNorm2d()
    (7): ReLU()
    (8): Dropout()
    (9): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (linear): Linear(in_features=16384, out_features=10, bias=True)
  (loss): CrossEntropyLoss()
)

```

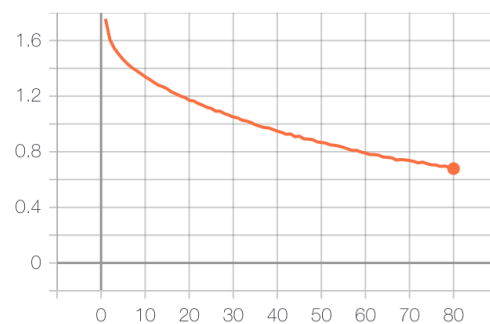
3.2 训练曲线

3.2.1 MLP

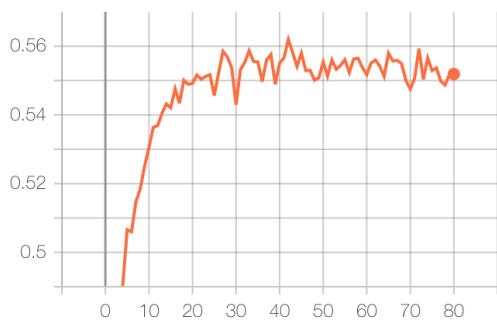
Accuracy
tag: Train/Accuracy



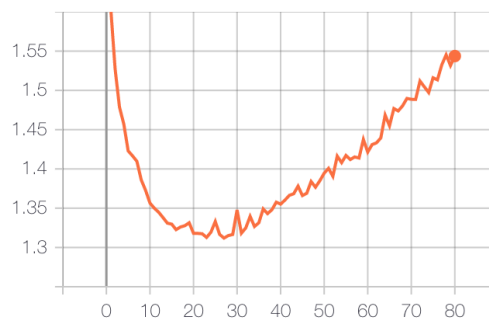
Loss
tag: Train/Loss



Accuracy
tag: Validation/Accuracy

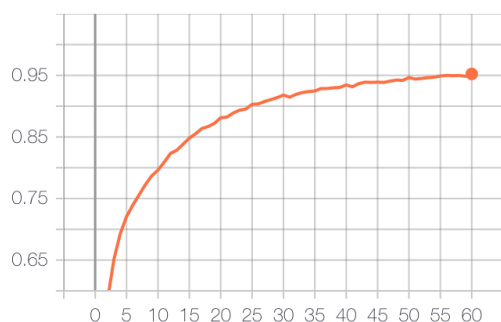


Loss
tag: Validation/Loss

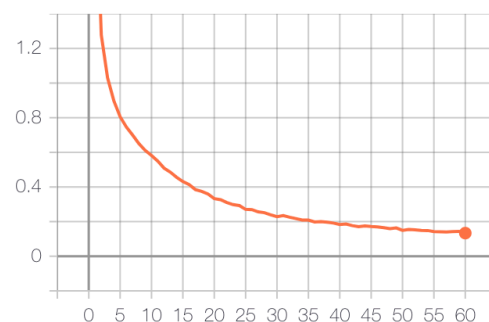


3.2.2 CNN

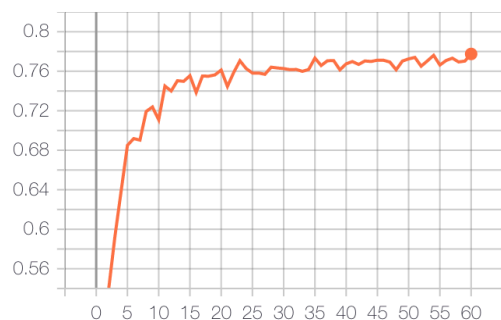
Accuracy
tag: Train/Accuracy



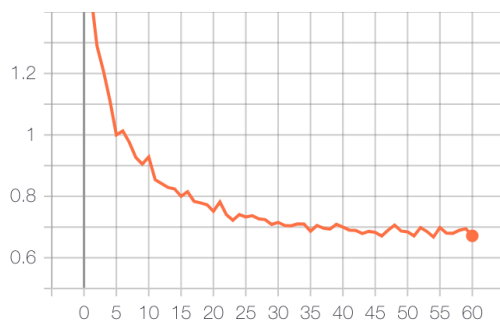
Loss
tag: Train/Loss



Accuracy
tag: Validation/Accuracy



Loss
tag: Validation/Loss



3.3 测试结果

	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
MLP	0.6786	0.7625	1.5437	0.5518	1.3506	0.5527
CNN	0.1335	0.9523	0.6714	0.7775	0.6817	0.7733

可以看出 CNN 比 MLP 的拟合能力更强，这是因为卷积层与池化层可以更好地捕获图像中的特征，CNN 的网络深度也更深。

4 思考题

1. Explain how self.training work. Why should training and testing be different?

这是一个用来表示模型当前状态的 bool 变量，如果是训练中，则 Dropout 会开启，batch normalization 也会进入训练模式，如果是训练完毕（eval 模式），则模型会直接输出计算结果。

2. Construct the MLP and CNN with batch normalization and dropout. Write down the hyperparameters that you use to obtain the best performance. Plot the loss value and accuracy (for both training and validation) against to every iteration during training.

模型参数、训练曲线见实验部分。

3. Explain why training loss and validation loss are different. How does the difference help you tuning hyper-parameters?

training 时 Dropout 是开启的，而且 batch normalization 是基于 training data 计算的平均值与方差，而 validation 时 Dropout 是关闭的，而且 batch normalization 是直接沿用的 training 阶段计算结果，在理想情况下（IID 假设下）validation data 的分布于 training data 相同，故无影响，但实际上分布会有一些偏差导致结果不同。

我们调参时应该尽可能增大模型的泛化能力，即在获得较高 training 准确率的同时关注 validation 准确率，如果相差过多则说明模型有 overfit 的风险。

4. Report the final accuracy for testing. Compare the differences between the results of MLP and CNN.

见实验结果部分。

5. Construct MLP and CNN without batch normalization, and discuss the effects of batch normalization.

以相同的参数对模型进行构造，含 batch normalization 的模型收敛要快得多，在相同的迭代次数下，batch normalization 可以有效地提高模型的准确率。

batch normalization 可以是梯度下降的方向计算更准确，可以认为在每一次计算梯度时都对数据进行了一定的白化，有效的避免了梯度消失等问题，在表现上提高了模型的泛化能力。

6. Tune the drop rate for MLP and CNN, respectively, and discuss the effects of dropout.

固定其他参数相同，实验表明 drop rate 在 0.5 左右可以获得最佳表现。

在神经网络中，dropout 率过低，模型容易过拟合，dropout 率过高，模型的拟合性能会过低而导致模型的表现变差。理论上 dropout 率在 0.5 时可以产生的等效网络个数最多，因此实际中也常常采用 dropout 参数为 0.5。

5 总结

本实验实现了两个不同的神经网络，并实现了 Dropout 与 Batch normalization 层，实验结果表明二者均对模型的训练有正面作用。