

ANN Homework2

实验概括

本次实验目标为使用多层感知机（MLP）或卷积网络（CNN）在 `Cifar-10` 数据集完成分类任务。

实验内容

`self.training` 的作用

在训练过程中，为了避免过拟合，会采用 `Dropout Layer` 随机drop一些节点/通道；在评估过程中不涉及参数更新，应当采用完整的网络进行评估；此外，在训练过程的 `BatchNormal Layer` 中，根据当前Batch的均值和方差进行标准化；在评估过程中，样本可能逐个被处理，此时无法在 `mini-batch` 中计算均值和方差，故采用训练过程中记录的 `running_mean` 和 `running_var` 做标准化操作。

通过 `self.training` 判断当前处于何种状态，对 `Dropout Layer` 和 `BatchNormal Layer` 进行调整。

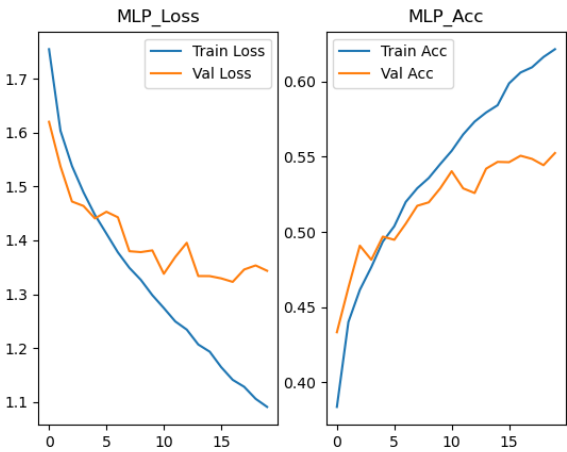
MLP&CNN 训练结果

1. MLP

1. Hyper - Parameters

Batch_size	Num_epochs	Learning_rate	Drop_date	Linear(1)	Linear(2)
100	20	1e-3	0.5	(3072,1024)	(1024,10)

2. Loss & Accuracy Curves



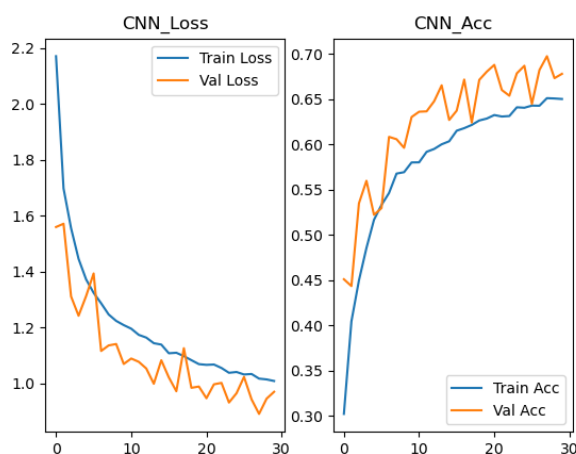
Test Accuracy: 54.7%

2. CNN

1. Hyper - Parameters

Batch_size	Num_epochs	Learning_rate	Drop_date	Conv2d(1)	Cov2d(2)
100	30	1e-3	0.5	(3,32,5,1,2)	(32,64,5,1,2)

2. Loss & Accuracy Curves



3. Test Accuracy: 69.2%

training_loss & validation_loss

training和validation对应不同的数据，训练过程中，模型会根据自身在training data上的表现更新参数拟合数据，而在validation data上并不会更新参数，因此理论上training_loss会低于validation_loss；根据模型在validation data上的表现，调整超参数。例如，当 $\text{training_loss} < \text{validation_loss}$ 时，倾向于认为模型处于under-fitting，可以考虑增加训练轮数；当 $\text{training_loss} > \text{validation_loss}$ 时，倾向于认为模型处于over-fitting，可以考虑增加Drop Layer、引入正则化项等。

Final Accuracy for Testing

MLP: 54.7% (作为后续比较的Benchmark)

CNN: 69.2% (Benchmark)

CNN的表现明显优于MLP，因为CNN在模型前期维持了图像的二维形态，可以学习到更准确的特征；而MLP在输入时就将图像展开成一维向量，图像中的位置信息丢失。

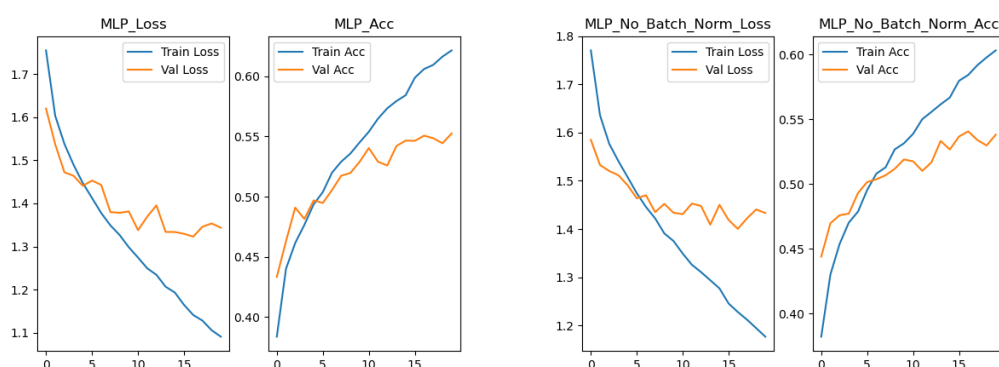
比较有无BatchNormal Layer的结果

1. MLP

1. 实验结果

有BatchNorm Layer: Test Acc: 54.7% (Benchmark)

无BatchNorm Layer: Test Acc: 53.4%



2. 结果分析

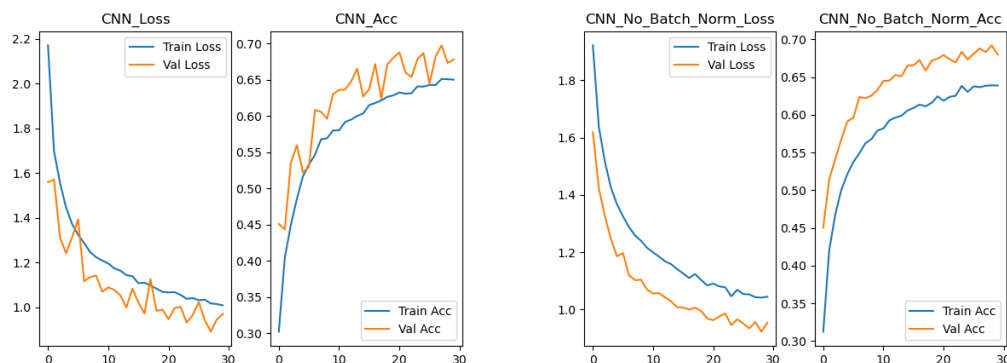
在多层感知机的训练过程中是否引入BatchNorm Layer并无明显差异，推测原因是MLP的输入中，一张图片的三个通道被全部压缩为一个一维向量，对来自不同通道的数据统一做归一化处理并不合理；

2. CNN

1. 实验结果

有BatchNormal Layer: Test Acc: 69.2% (Benchmark)

无BatchNormal Layer: Test Acc: 69.1%



2. 结果分析

理论上，BatchNorm Layer可以加速模型收敛，因为数据已经被归一化处理到相同的分布，符合独立同分布假设；但是在实验中，去掉BatchNorm Layer后模型的收敛性反而更好。

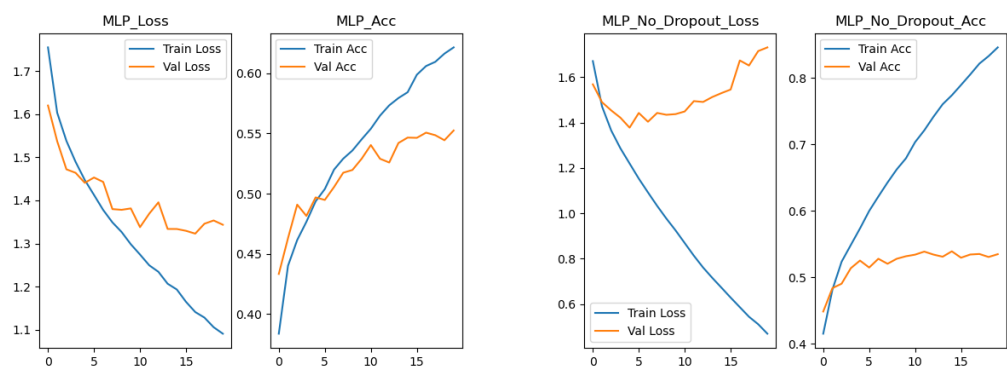
比较有无Dropout Layer的结果

1. MLP

1. 实验结果

有Dropout Layer: Test Acc: 54.7% (Benchmark)

无Dropout Layer: Test Acc: 52.9%



2. 结果分析

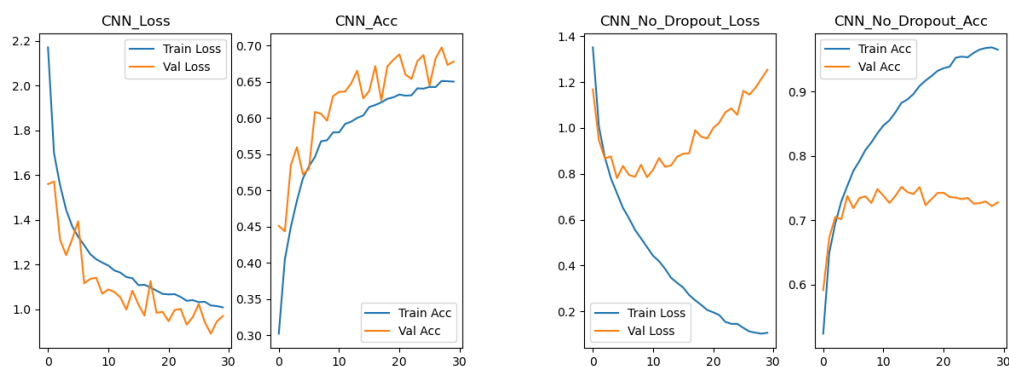
取消Dropout Layer后，模型发生了明显的过拟合，在第5个epoch后validation loss就不再下降，可以明显看出Dropout Layer防止过拟合的作用。

2. CNN

1. 实验结果

有Dropout Layer: Test Acc: 69.2% (Benchmark)

无Dropout Layer: Test Acc: 73.9%



2. 结果分析

1. 去掉Dropout Layer后, 模型产生了明显的过拟合, 在第10个epoch之后, 模型在validation data上的损失不降反增, 而在test data上的损失持续下降, 最终在训练集上的准确率达到了96.5%, 明显过拟合;
2. 去掉Dropout Layer后虽然产生了过拟合, 但是最终的Test Accuracy高于加入了Dropout Layer的原模型。推测原因是, 原模型 (benchmark) 设置的dropout rate为0.5, 较大, 以至于模型不能充分学习数据特征, 因此准确率较低。

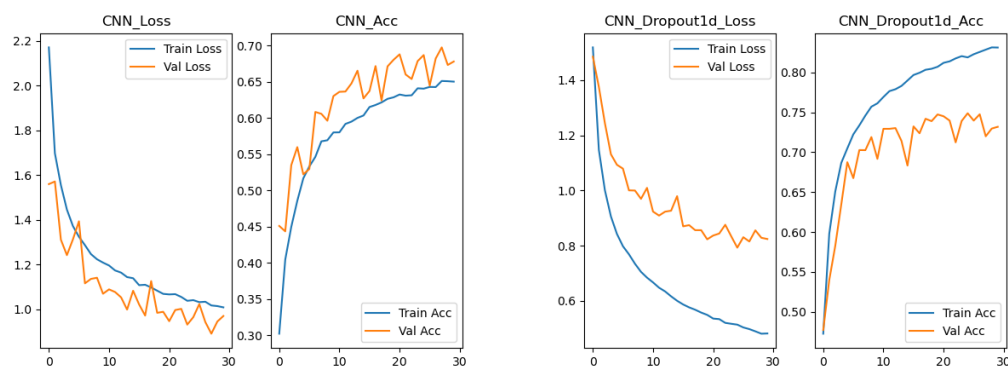
附加问题

Dropout1d

1. 实验结果

Dropout2d: Test Acc: 69.2% (Benchmark)

Dropout1d: Test Acc: 75.2%



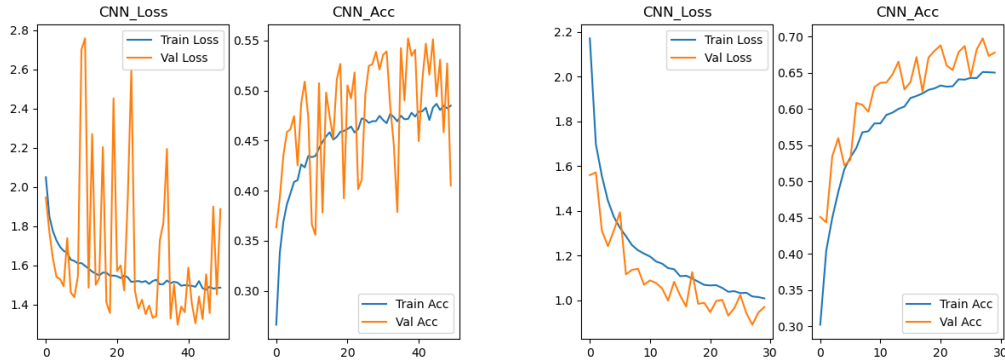
2. 结果分析

1. Dropout1d不能有效避免过拟合, 可以看出在第10个epoch后出现了过拟合现象;
2. 虽然Test Acc有所提升, 但Benchmark调整Dropout Rate后也能达到近似准确率 (见实验报告最后一小节)。

Fine-tune

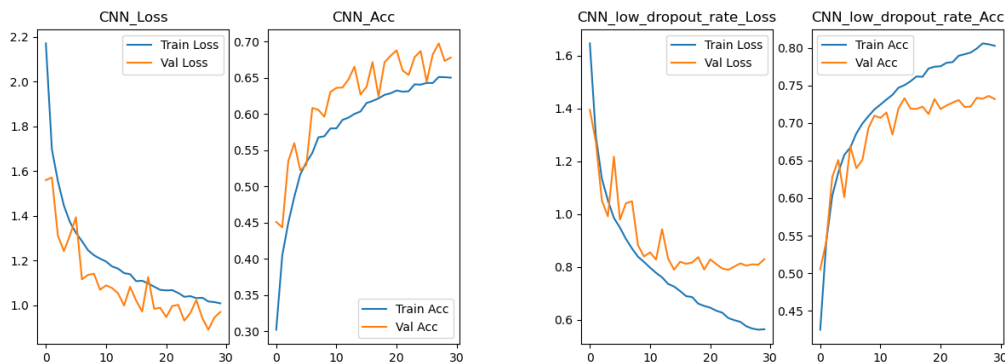
1. 卷积层的通道数设置

在最初的实验中，两个Cov2d的input_channel, output_channel分别为 (3, 8) , (8, 16) , 模型无法在训练中收敛，波动性非常大；将上述参数替换为 (3, 32) , (32, 64) 后模型的收敛性明显改善，并且准确率更高。推测原因是更多的通道可以学习输入向量的更多特征。



2. Dropout_rate 设置

由前文的分析可以看出，较大的dropout_rate使得模型不能充分学习到输入数据的特征，因此尝试将 dropout_rate从0.5调整至0.2，结果明显提高，Test Acc从69.2% (Benchmark) 提升至72.7%；但与此同时，出现了一定的过拟合现象。



3. Batch_size 的设置

将 Batch_size 从100调整至50后，模型的准确率有所提高；

Batch_size=100; Drop_rate=0.2 : Test Acc: 72.7%

Batch_size=50; Drop_rate=0.2 : Test Acc: 74.3%

推测原因是更小的Batch使得模型在每个Batch更新模型参数的过程中更易于进行拟合，因此准确率上升。

