HW2

计 81 卢星宇 2018011280

self.training 的作用

在本次实验中，self.training 的作用主要是决定是否启用 BatchNorm 和 Dropout，因为在 train 时，self.training 为 True，而在 evaluate 时 self.training 为 Fasle。之所以 test 和 train 应该区别对待，是因为不能利用测试集更改模型参数，否则训练集和测试集就不存在区别，实验结果也不具有任何意义。

超参数调整

MLP

主要针对 learningrate 和 droprate 两个参数进行调整
模型结构：

```
(0): Linear(in_features=3072, out_features=200, bias=True)
(1): BatchNorm1d()
(2): ReLU()
(3): Dropout()
(4): Linear(in_features=200, out_features=10, bias=True)
```

对 learning rate 做出调整，获得结果如下

rate	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007
test_acc	52.82%	53.68%	53.64%	54.25%	54.03%	53.56%	53.61%
0.0008	0.0009	0.001	0.002	0.003	0.004	0.005	0.006
53.63%	54.16%	54.51%	53.20%	53.36%	53.40%	53.12%	52.69%

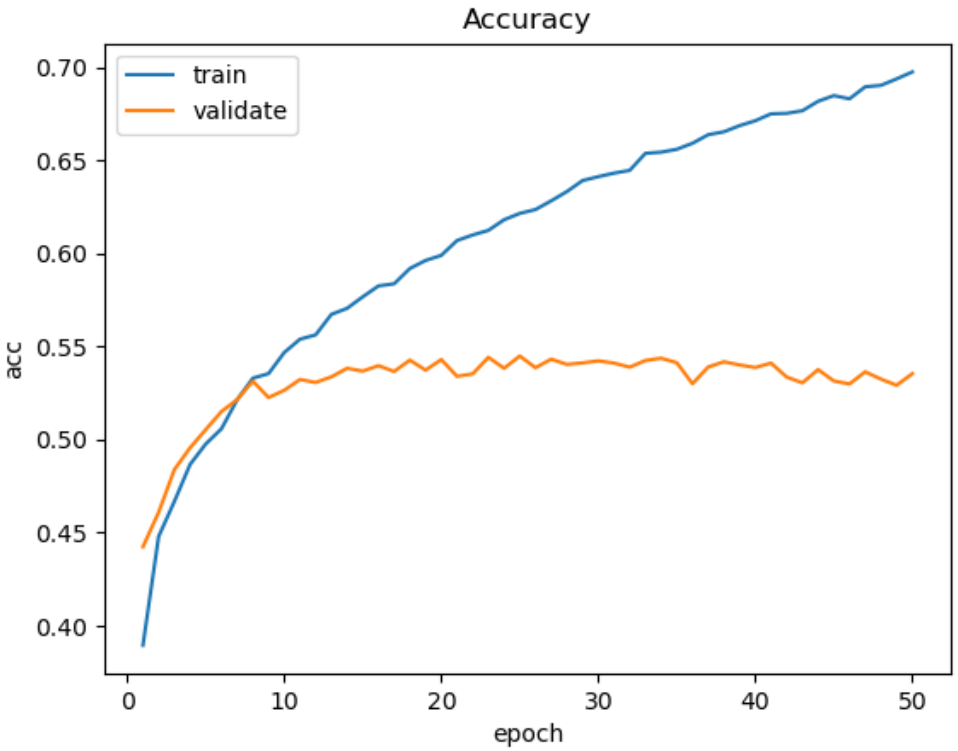
所以我认为本实验最佳的 learning rate 即为 0.001
然后在 learning rate = 0.001 的情况下探索 drop_rate 的最佳值

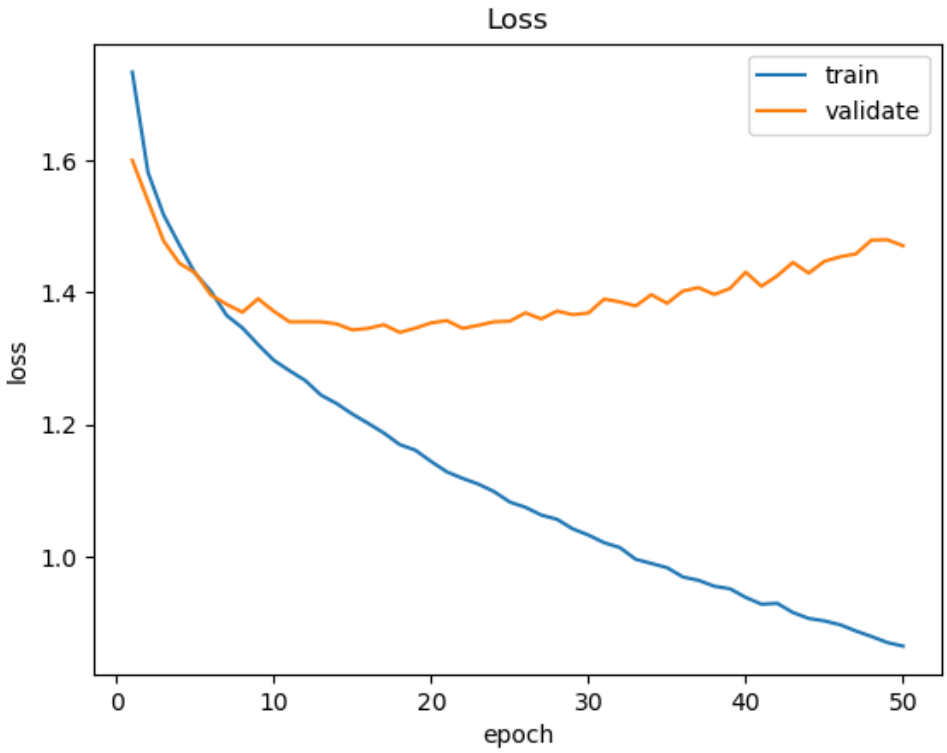
rate	0.05	0.06	0.07	0.08	0.09
test_acc	53.39%	53.31%	53.62%	52.82%	53.40%
0.1	0.2	0.3	0.4	0.5	0.6
54.22%	52.99%	54.04%	53.88%	53.42%	53.04%

可以看出对于 MLP，dropout rate 的影响随机性很大，以 0.3 为最终的准确率。
按照最优的超参数

```
('--batch_size', type=int, default=100)
('--num_epochs', type=int, default=50)
('--learning_rate', type=float, default=1e-3)
('--drop_rate', type=float, default=0.3)
```

得到的实验结果图像如下





CNN

网络结构

```
(layers): Sequential(
  (0): Conv2d(3, 100, kernel_size=(5, 5), stride=(1, 1))
  (1): BatchNorm2d()
  (2): ReLU()
  (3): Dropout()
  (4): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (5): Conv2d(100, 50, kernel_size=(3, 3), stride=(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=1800, out_features=10, bias=True)
```

对 learning rate 做出调整，获得结果如下

rate	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008
test_acc	74.64%	74.98%	75.28%	75.72%	75.97%	76.45%	77.58%
0.0009	0.001	0.0011	0.0012	0.0013	0.0014	0.0015	0.0016
76.36%	76.26%	76.78%	76.42%	76.58%	76.70%	75.45%	75.82%

所以我认为本实验最佳的 learning rate 即为 0.0008

然后在 learning rate = 0.0008 的情况下探索 drop_rate 的最佳值

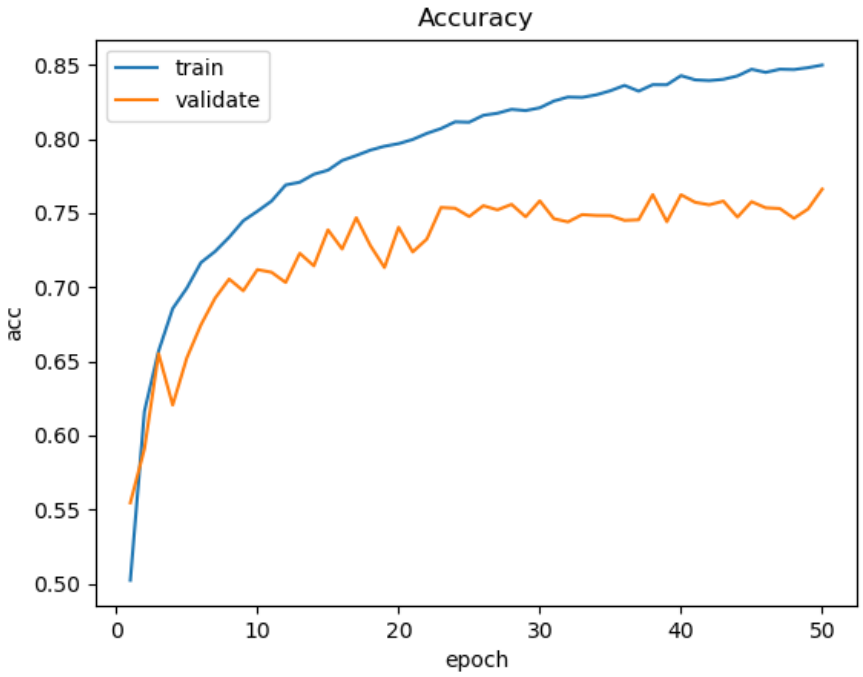
rate	0.1	0.2	0.3	0.4
test_acc	74.04%	75.21%	75.82%	76.59%
0.5	0.6	0.7	0.8	0.9
75.62%	75.98%	75.29%	72.42%	60.57%

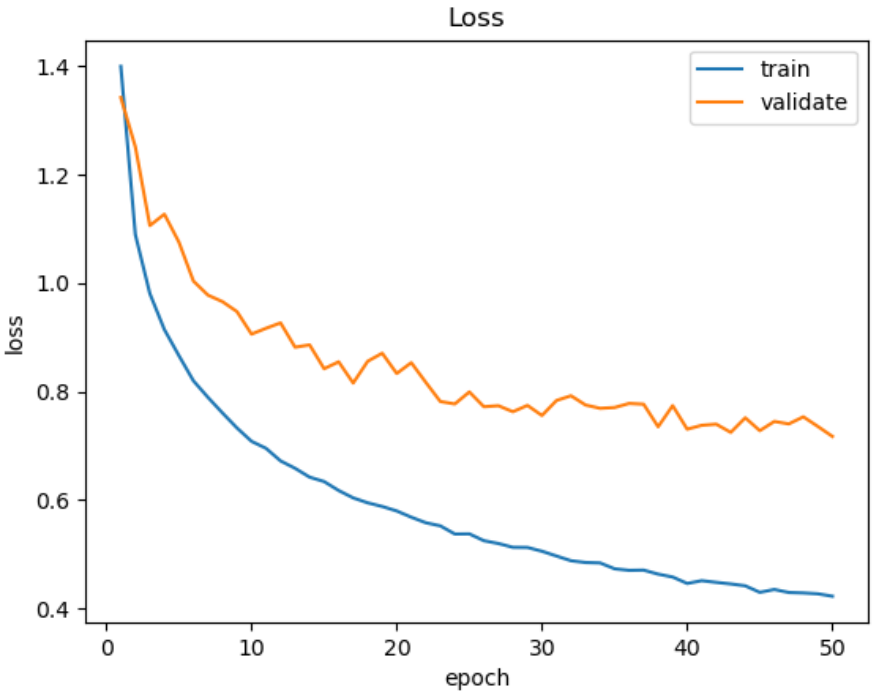
我认为在本实验中最佳的 drop_rate 为 0.4

最终超参数设置：

```
'--batch_size', type=int, default=100)
'--num_epochs', type=int, default=50)
'--learning_rate', type=float, default=8e-4)
'--drop_rate', type=float, default=0.4)
```

实验结果


Accuracy
train
validate
acc
epoch
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0
10
20
30
40
50


Loss
train
validate
loss
epoch
1.4
1.2
1.0
0.8
0.6
0.4
0
10
20
30
40
50

为什么训练集和验证集 loss 不同，如何利用这种不同调整超参数

答：同样参数在两种集上 loss 不同的根本原因是两种集中的数据分布不同，而由于模型是根据训练集调整参数的，故而出现了过拟合现象，即参数中包含了训练集上数据分布的某种特征而验证集上不存在这种特征，所以 loss 出现了不同。

这种不同的主要特征是，验证集的 loss 先降后升，这种现象说明了每一时刻的过拟合现象严重程度如何。促使我们采取提高泛化性能的措施，如提高 drop_Rate，使用 batchNorm 等。

报告测试集上的准确率，对比 MLP 和 CNN 上的表现，简要说明原因

答：在测试集上 MLP 和 CNN 分别取得了 54.51%和 77.58%的准确率。这种原因主要是模型结构不同导致的。

MLP 的缺点在于，所有的神经元之间紧密连接，因此当图片大小和类别增大时，参数量会迅速增大，训练速度变慢，且很容易过拟合，泛化能力很差。主要表现在 MLP 不具有平移不变性，比如如果训练集中某一种物体集中在图片的某一位置，那么当测试集中该种物体出现在另一位置时，就很难给出正确的结果。而且 MLP 无法利用图片的局部信息，因为在 MLP 中，一个像素与其他像素之间被排列成 1 维，会丢失二维上的信息。

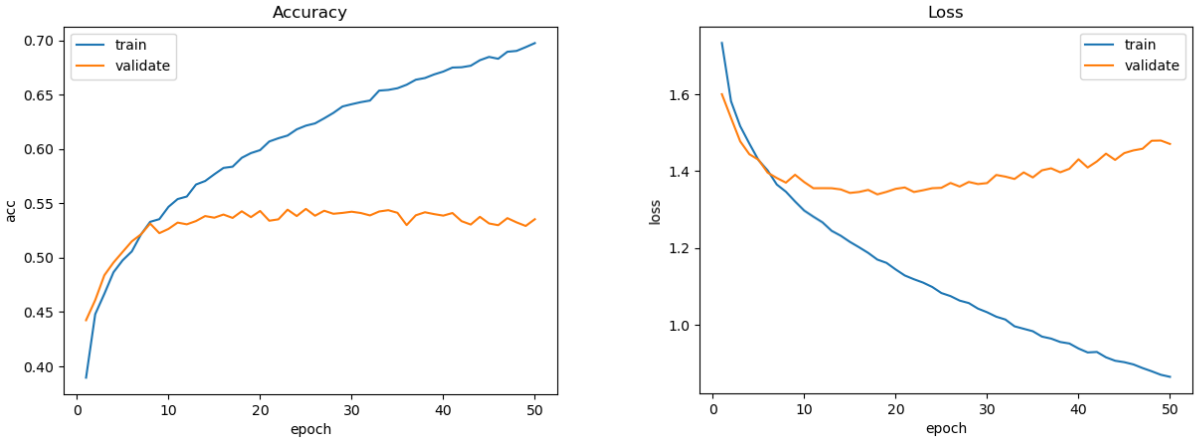
而 CNN 针对这些做出了改进，首先不同位置的卷积核之间共享参数，这就改正了 MLP 不具有平移不变性的缺点，且参数都是在二维层面上提取特征，可以利用图片局部信息，丢失的信息会减少很多。

不使用 BatchNorm，对比两种模型的效果，说明 BatchNorm 的作用

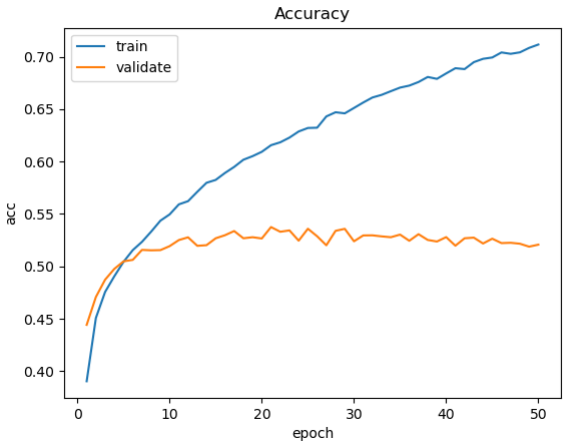
其余超参数均为上述实验中的最佳值

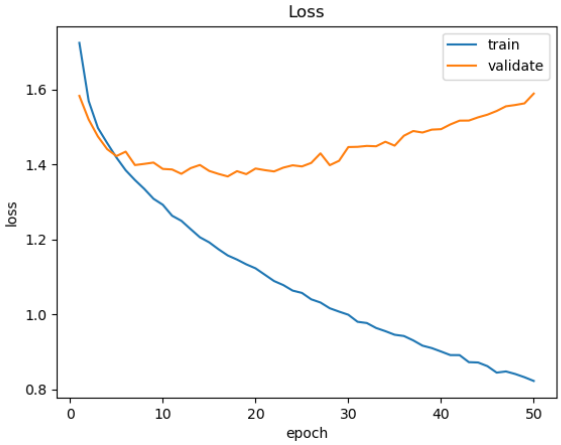
MLP

使用 BatchNorm



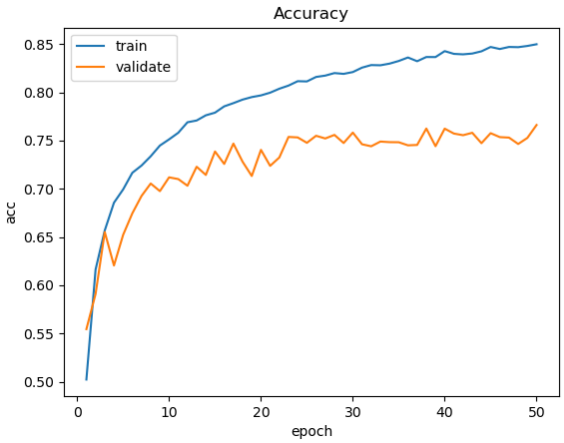
不使用 BatchNorm

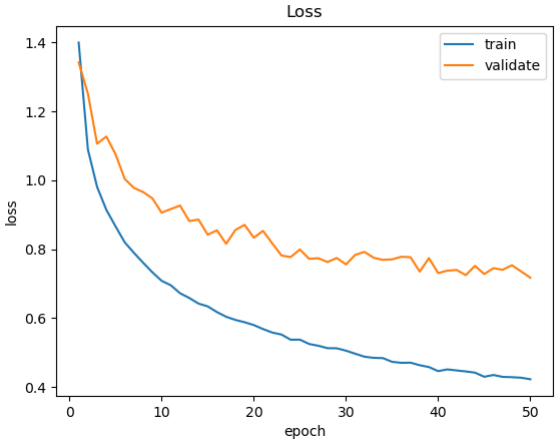




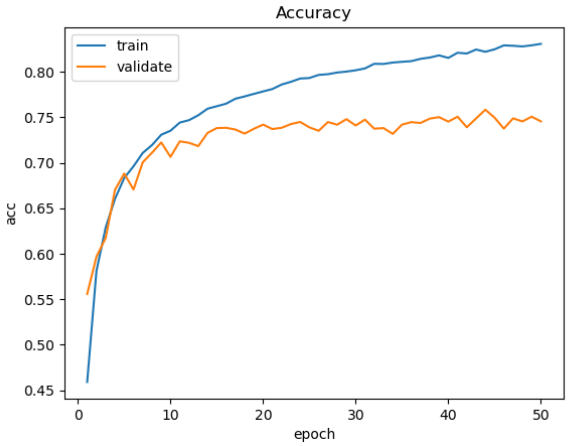
CNN

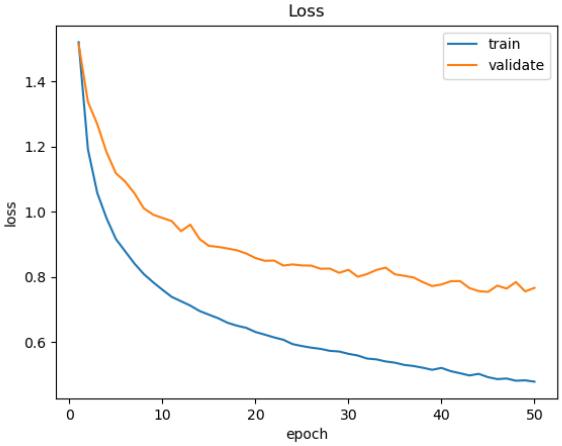
使用 BatchNorm





不使用 BatchNorm





可以看出，在使用了 BatchNorm 之后，过拟合现象得到了缓解，验证集上的 loss 在过拟合发生之后上升的更加缓慢，取得的最佳 accuracy 也更高。

BatchNorm 的作用

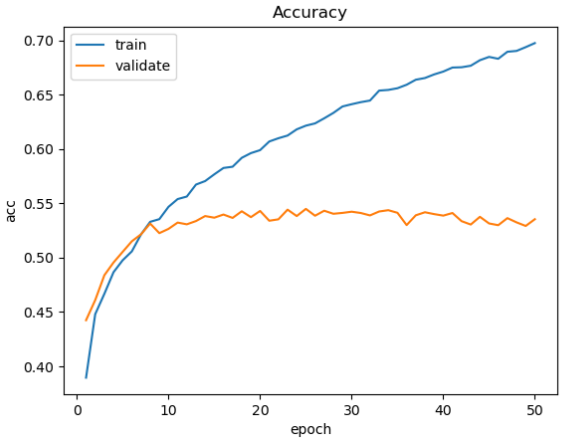
1.加快收敛速度，有效避免梯度消失。
2.提升模型泛化能力，BN 的缩放因子可以有效的识别对网络贡献不大的神经元，经过激活函数后可以自动削弱或消除一些神经元。另外，由于归一化，很少发生数据分布不同导致的参数变动过大问题。

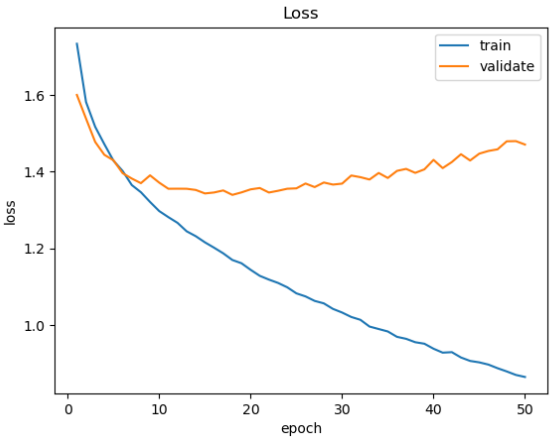
不使用 Dropout，对比两种模型的效果，说明 Dropout 作用

其余超参数均为上述实验中的最佳值

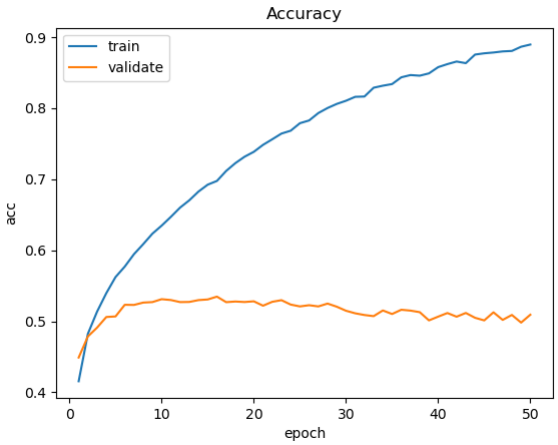
MLP

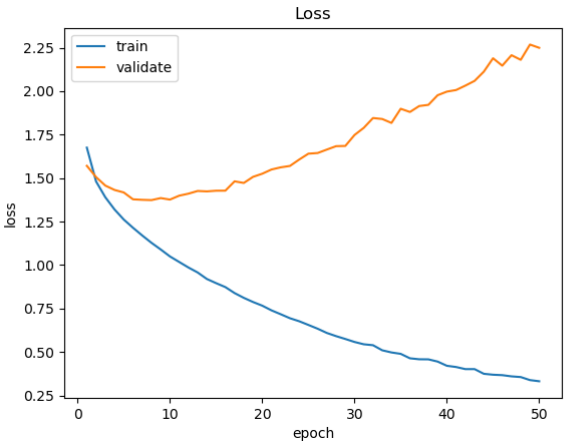
使用 Dropout





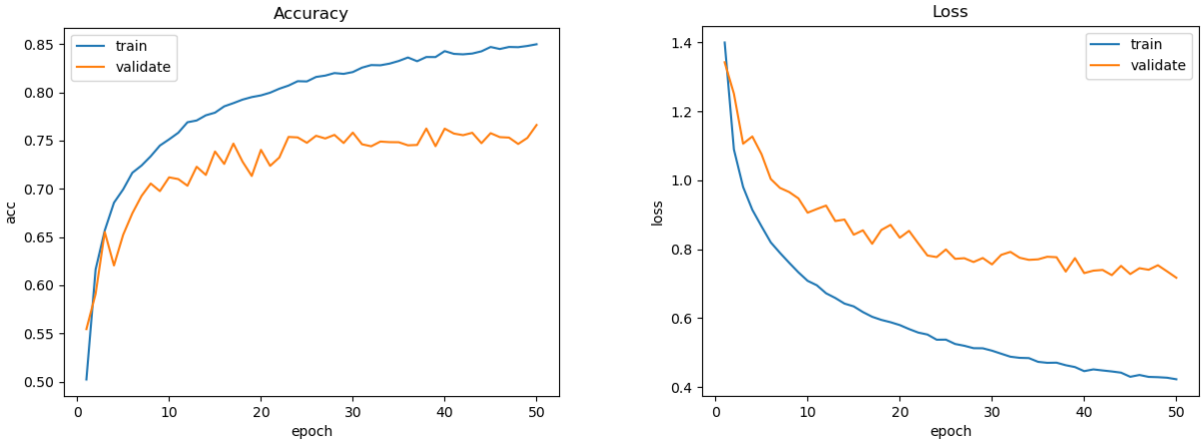
不使用 Dropout



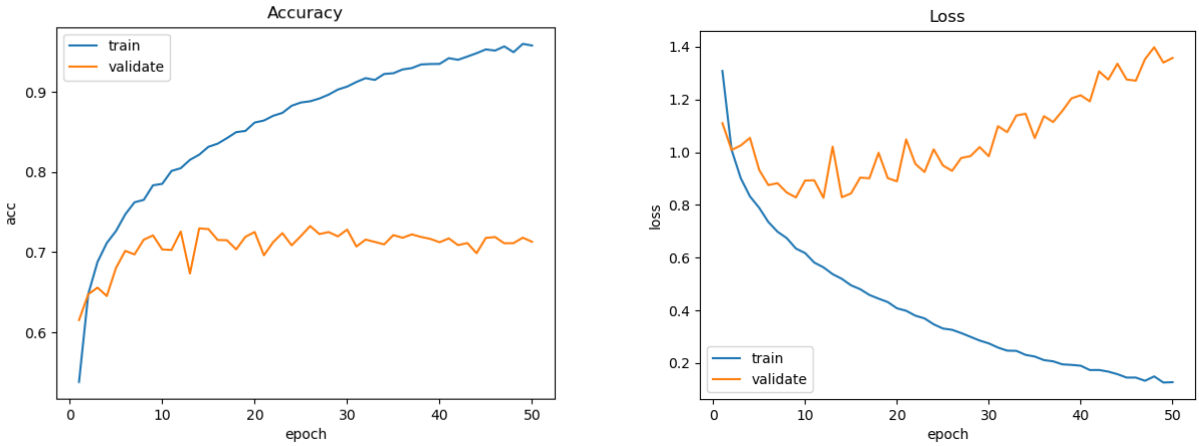


CNN

使用 Dropout



不使用 Dropout



可以看出，在使用了 dropout 之后，大大缓解了过拟合现象，验证集上的 loss 在过拟合发生之后上升的更加缓慢，取得的最佳 accuracy 也更高。

dropout 的作用

1. dropout 导致两个神经元不一定每次都同时出现。这样权值的更新不再依赖于有固定关系的隐含节点的共同作用，阻止了某些特征仅仅在其它特定特征下才有效果的情况 。迫使网络去学习更加鲁棒的特征

2. 将 dropout 比作是有性繁殖，将基因随机进行拆分，可以将优秀的基因传下来，并且降低基因之间的联合适应性，使得复杂的大段大段基因联合适应性变成比较小的一个一个小段基因的联合适应性。