

# 神经网络 第二次作业报告

2017012289 李岱轩

## 1 实验内容与附加实现

本次实验通过实现带有Dropout和BatchNormalization的CNN和MLP，并进一步实验验证其效果，来学习CNN与MLP以及相关优化技术的特征。

除了基本的实现之外，我在实现中，为BatchNorm增加了momentum选项来进一步提高计算moving average的灵活性，momentum的默认值设定为0.1，因此可能会与机判结果不同。

注：在上一次作业中，我在欧几里得Loss之前加入了softmax来增加准确性，而导致机判结果与直接欧几里得Loss不同，希望能进行申诉。

## 2 实验结果与问题回答

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

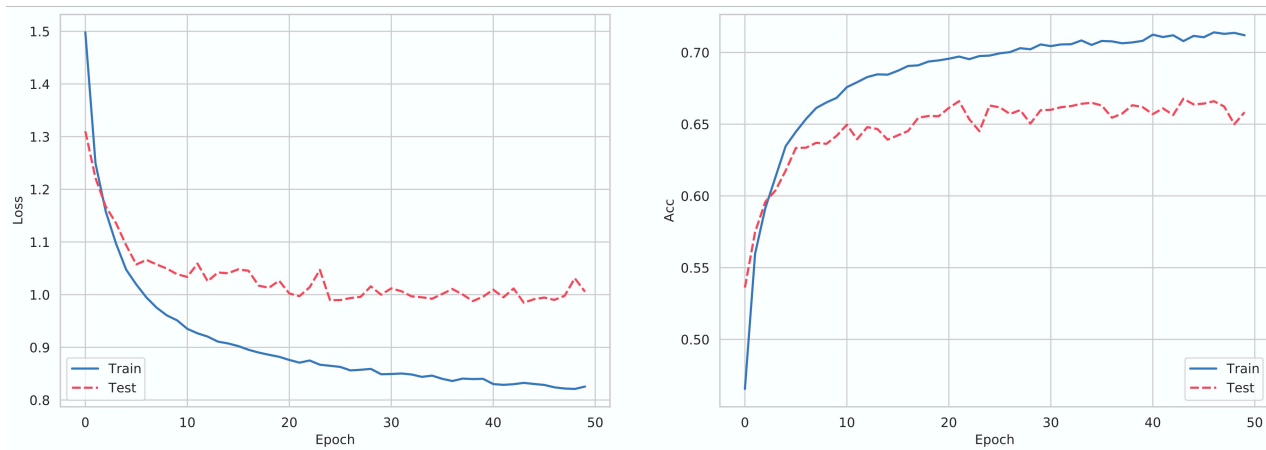
当进行model.eval时，self.training自动设置为False，从而达到区分测试和验证集的效果。需要这样做，是因为在训练集中我们增量计算moving average，测试集不能计算（因为不能干扰训练集），而且测试集要利用训练集计算的moving average作为样本均值和方差。在编程实现上的不同，导致需要self.training这样一个值来进行区分。

### 2.2 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.

CNN最好结果：

Batch_Size	Use_BatchNorm	Momentum	Drop_Rate	Learning_Rate	Filter Num(中间层)
100	True	0.1	0.1	1e-3	10; 20

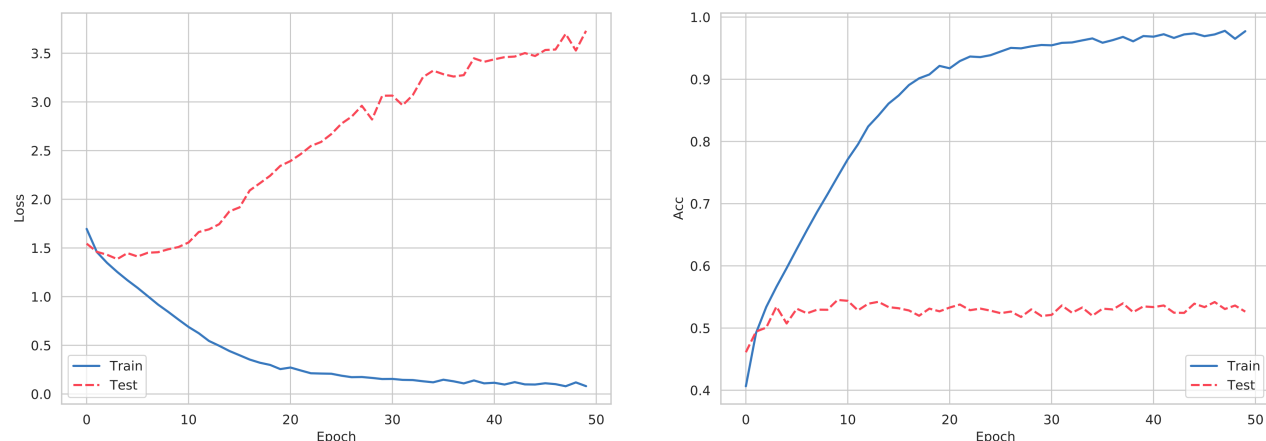
最高准确率为66.76%，Loss与Acc如下图所示。



### MLP最好结果:

Batch_Size	Use_BatchNorm	Momentum	Drop_Rate	Learning_Rate	Hidden Layer
100	True	0.1	0	1e-3	1536

最高准确率为55.09%, Loss与Acc如下图所示。



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

模型在训练集上进行梯度下降，而只在验证集上进行验证，因此两者不同。训练集上训练时间过长，Loss会不断下降而会出现过拟合，不能很好的预测新的数据。非过拟合情况下，验证集和测试集几乎同分布，因此通过观察验证集趋势可以辅助调参，比如可以适当提高Drop rate防止过拟合，或者降低 learning rate以防止震荡。

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

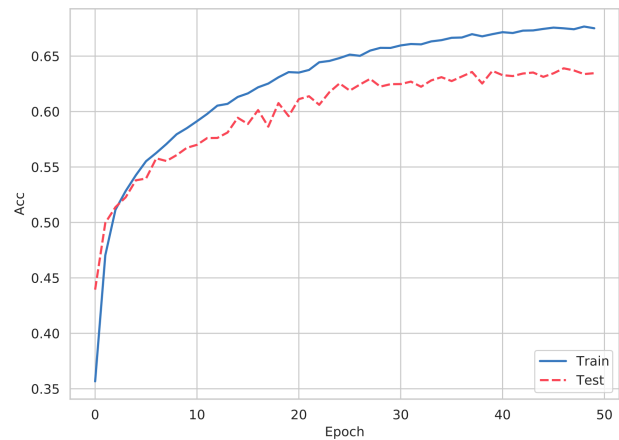
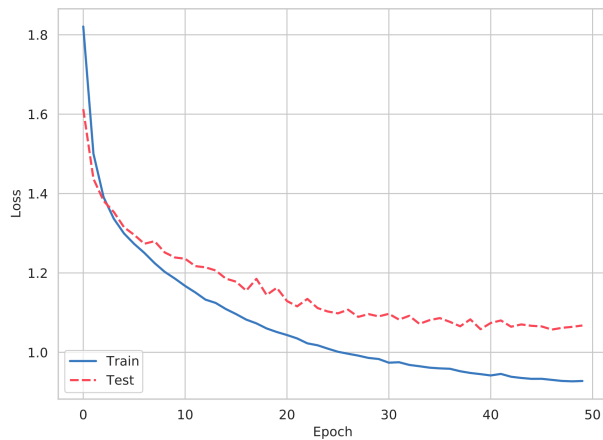
全部实验中最好准确率结果为66.76%，使用CNN的超参数在3中已经进行展示。

比较上面两张图可以得出结论，CNN在训练上更加鲁棒，波动较小，收敛较快，而且不容易发生过拟合。

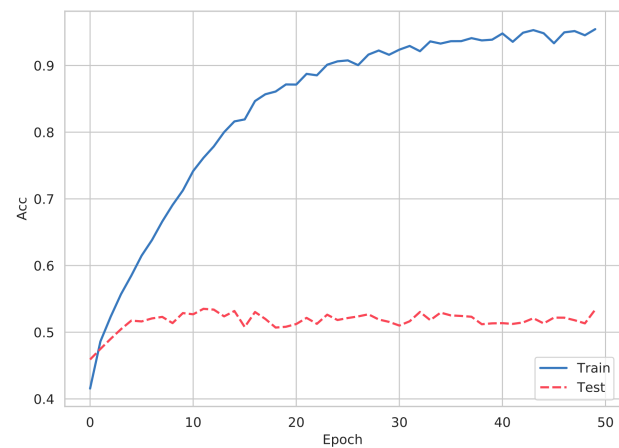
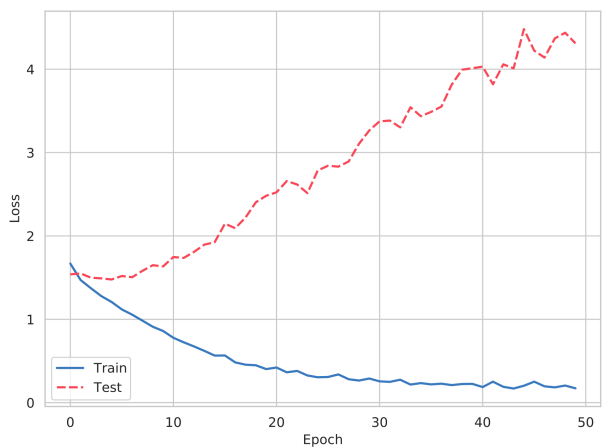
## 5 Construct MLP and CNN without batch normalization, and discuss the effects of batch normalization

分别在CNN和MLP最高准确率的位置，只去掉BatchNorm进行实验，得到的结果如下图所示。

CNN结果，最高准确率为63.89%。



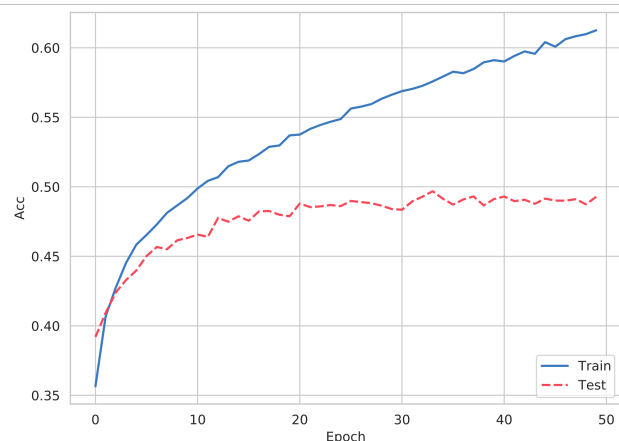
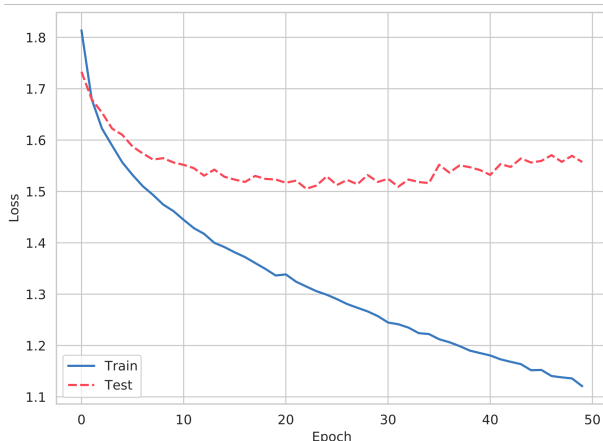
MLP结果，最高准确率为54.79%。



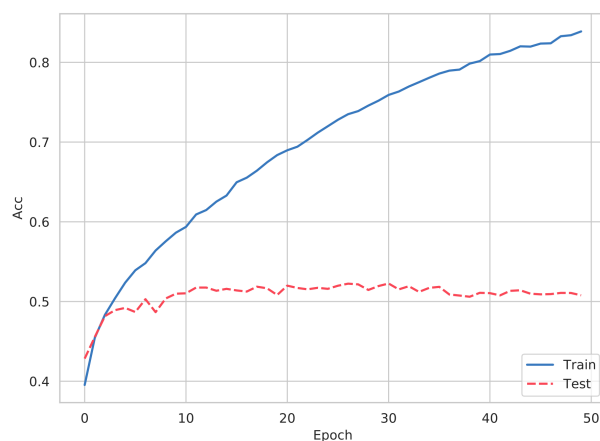
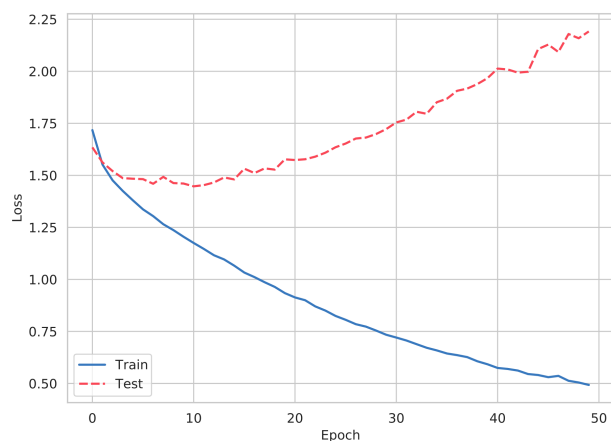
可以看出，增加了BatchNorm的神经网络，收敛更快，更不容易发生过拟合，波动较小，最终最高准确率也较高，性能较好。

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

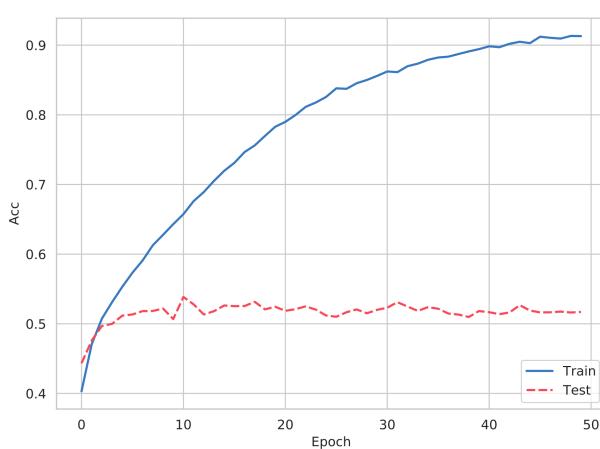
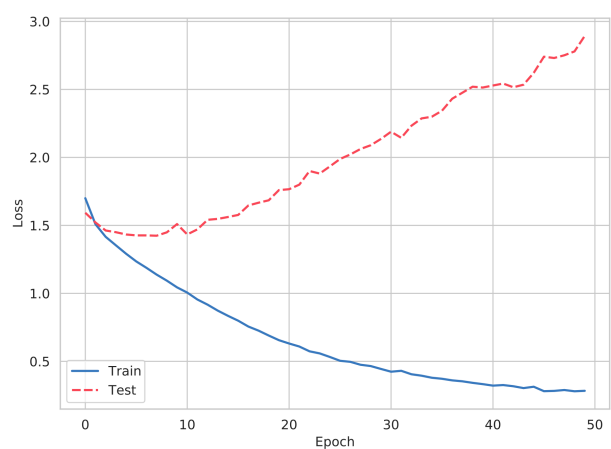
以下三张图为MLP在Drop rate为0.8， 0.5， 0.3时的准确率与Loss。



Drop Rate = 0.8, 最高准确率49.3%。

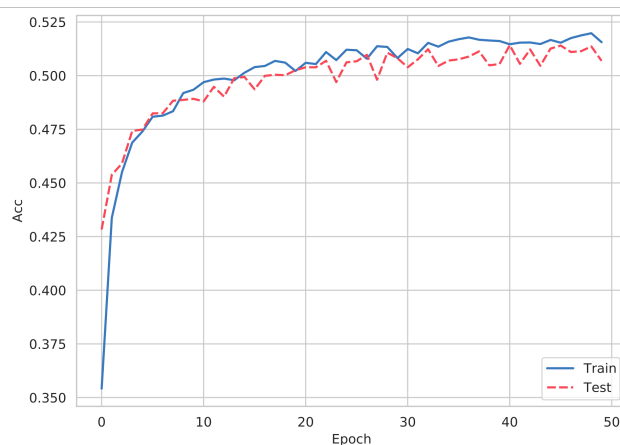
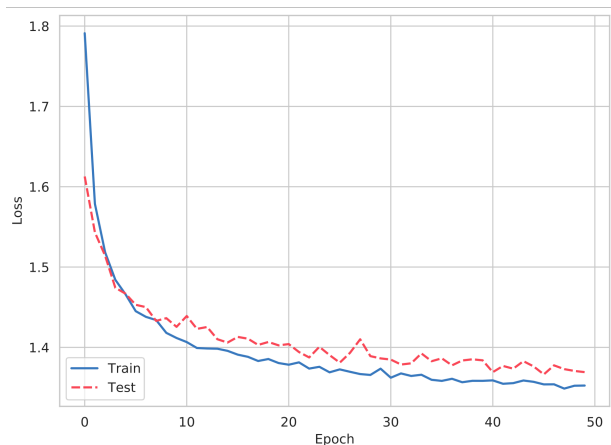


Drop Rate = 0.5, 最高准确率52.4%。

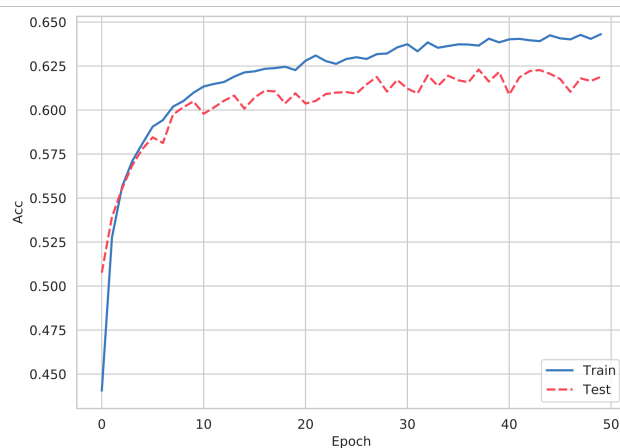
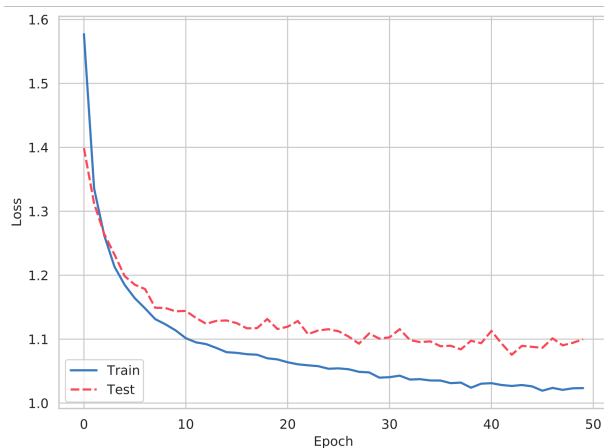


Drop Rate = 0.3, 最高准确率53.2%。

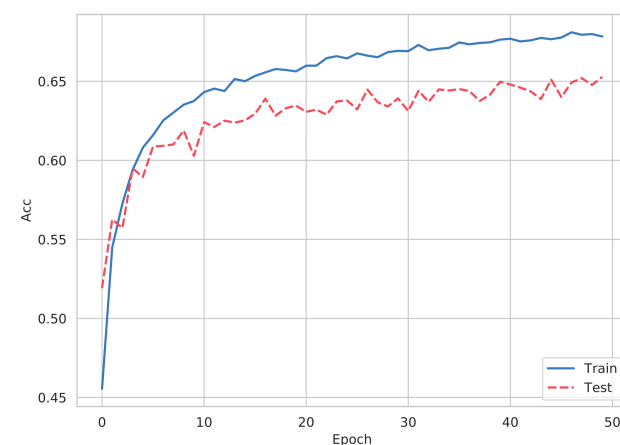
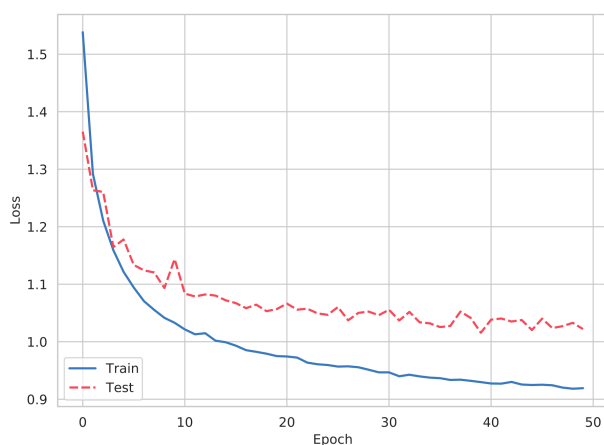
以下三张图为CNN在Drop rate为0.8, 0.5, 0.3时的准确率与Loss。



Drop Rate = 0.8, 最高准确率51.7%。



Drop Rate = 0.5, 最高准确率62.3%。



Drop Rate = 0.3, 最高准确率65.4%。

可以得出以下结论：

- Drop Rate越高，曲线越容易过拟合，反之则不容易出现过拟合的现象；
- Drop Rate越高，丢失的信息越多，波动越大，且Drop Rate过大会严重影响模型性能；
- Dropout层应该看具体情况选择性添加。例如本实验中，MLP和CNN规模较小，发生过拟合风险很小，可以不必添加DropOut（尤其是MLP中）。实验中明显观察到Dropout很小的时候性能更好。