

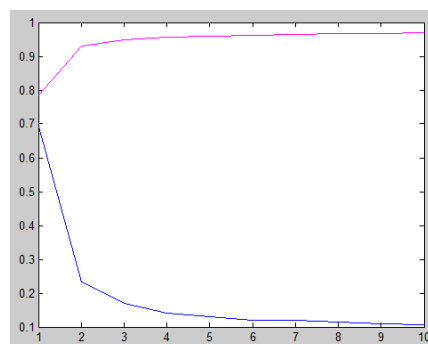
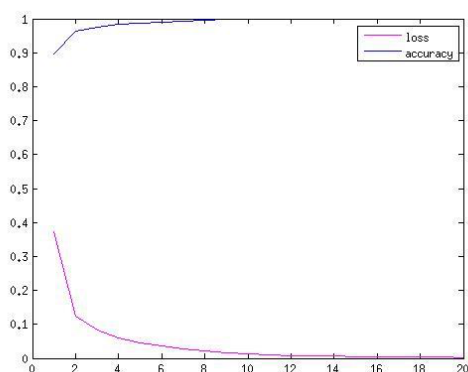
3rd Report of Artificial neural network

2013012245 白可

第三次实验实现了一个简单的 CNN 网络，具有两个卷积层，两个 pooling 层。图片为 28*28 的数字图片。由于图片数目大，且运算时需要采用“循环遍历”等操作，所以实际实验中，真正使用全部训练集（六万张）和测试集（一万张）的总共进行了一次实验。其余实验分别使用六千张训练集和一千张测试集。取得了与较大的数据集相似的效果。以下是我在实验中得出的结论：

一、CNN 网络在数字分类上的大量样本上没有体现出明显的优越性。

但是它需要的参数更少



左图为 Relu + sigmoid 的两层网络，右侧为实验中的 CNN 网络，可以看到，Relu+sigmoid 网络无论是在收敛速度上还是准确率上都明显优于 CNN。

	训练准确度	测试准确度	训练损失	测试损失	速度
CNN	96.90%	97.34%	0.10678	0.088	2h40min
Softmax	100.00%	98.15%	0.0027	0.0687	84s

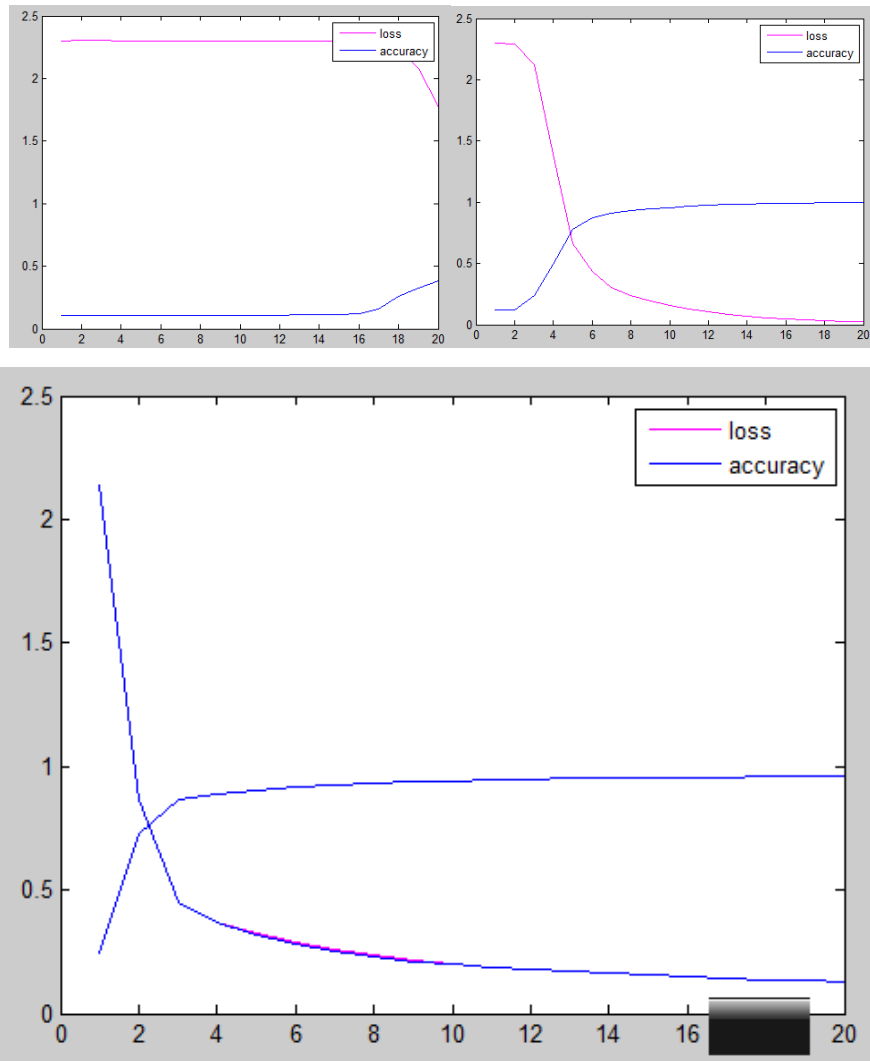
查了相关的资料暂时没有发现合理的解释。但我认为，在 CNN 里，我们采用了“权值共享”这一机制，整个的参数一共有 $9*8 + 9*4 + 196*10$ 个，但是一个两层的 MPL，一共需要 $784*256*10$ 个参数，显然，后者的参数更多，多层的感知机理论上可能实现任何的分类。但是一旦图片复杂，它就无法胜任了。

二、CNN 在较小的数据集上性能优于 MPL

我使用了 6000 的训练集和 1000 的测试集，分别对 sigmoid/relu/CNN +softmaxlayer 进行了测试：

	训练准确度	测试准确度	训练损失	测试损失	速度
CNN	96.23%	94.20%	0.13	0.18	34min
Relu	99.72%	92.20%	0.025	0.2518	13s
Sigmoid	38.25%	41.6%	1.73	1.64	30s

Sigmoid 暂时完败，



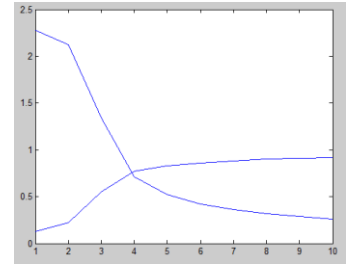
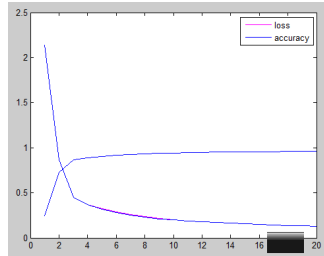
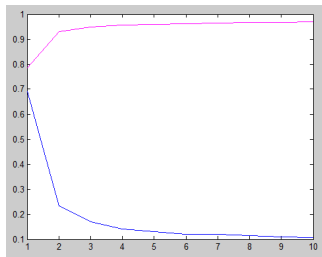
左上 sigmoid, 右上 Relu, 下面是 CNN，可以看到，它的收敛速度很快。我想，大致也是因为它参数比较少，所以训练好了基本就不会发生变化了。

我觉得这个实验结果很有启发性，

少量数据集就能达到比较好的效果。开始的时候我把迭代次数设成了 10，后来发现在后来阶段，它正确率依然以 0.5 的速度上涨，于是变调整成了 20，得到不错的效果。那么究竟多少数据就能有比较好的效果呢？

三、数据集大小与准确率之间的关系。

	训练准确度	测试准确度	训练损失	测试损失	速度
60000*10	96.23%	94.20%	0.13	0.18	34min
6000*20	96.90%	97.34%	0.10678	0.088	2h40min
3000*10	91.3%	93.6%	0.27	0.32	9min
1000*10	85.9%	84%	0.487	0.42	3min



可见，准确度还是和数据集有一定的相关度。但是在参数较少的时候，可能并不需要如此大量的数据集。

数据集较小时，通过增添训练的次数，可以提升准确率。（表格第一行与第二行可以看到对比）

四、第一层的 filter 图像



由于保存成 jpg 格式的图片由于编码的不同，原图信息消失。采用截图方式，边上的灰色边缘是截图过程中的背景。观察这些 filter，没有发现他们与“数字”类似的特征，可能是因为他们比较小，因此无法反映出原图的特征。

总结：

这次实验，体会到了“权值共享”这一特点。整个流程的关键之一在于 back propagation 的过程，不断将权值向我们希望的地方调整。想想之前的 MPL 其实也是同样的思路。但在大规模网络的训练上 CNN 效果却十分好。值得未来更多的探索。