本文档对网络调试的方法进行说明，**详细的调试的过程见 log.pdf 文件的内容**（记录了在调试过程中不同参数产生的不同结果）。

1.总体说明：

在 BP 网的实现中，主要用了两种不同的实现方式，在 BPNN 包下。这两种实现方式在网络的结构框架上是相同的，区别在于 SigmoidBPNN 的激活函数采用的为 sigmoid 函数，TanhBPNN 采用了 tanh（导数为 $1-f^2$）函数，相应的网络权重的调整算法也有所区别，但思想都是梯度下降法。

对 CNN 的实现中，调用了 keras 框架和 tensorflow 工具，使得工作量减少很多，只需理解卷积网络的结构即可正确使用。

2.LcdD 识别

这一部分开始采用了 SigmoidBPNN 的实现，因为输入和输出取值都为 0、1，使用 sigmoid 函数作为非线性函数是合适的，因为 sigmoid 函数值域在（0,1）间。但是考虑到 tanh 函数能够更好地收敛，所以将输入输出值处理为-1 和 1。

3.Sine 拟合

开始时，也采用了 SigmoidBPNN 实现，发现在训练结束后，通过命令行输入产生输出，发现当输入为正值时，预测的结果误差勉强能接受，但当输入为负值时，基本上预测的输出都是一个很小的接近零正数。调整几次后，发现仍然有这个问题，再次仔细思考该网络的建立时发现了问题，所采用的 sigmoid 函数的值域为（0,1），故不管如何去调整都不可能得到理想的结果。

为了解决上面的问题，开始想到的办法是在训练时将 Sine 函数的值做一个变换，使其落在[0,1]，在验证的时候，将产生的预测结果再做一个逆变换还原到真实的值。但该方法在调整训练了几次后发现效果也不是很好，误差很大。（分析见**注**）

查阅书籍后发现在构建 BP 网的过程中，激活函数的选取不仅仅有 sigmoid 函数，还有一个使用较多的函数为 tanh 函数，该函数相较于 sigmoid 函数优势在于值域为（0,1），建模的范围有很大的提升。同时该函数的求导比较简单，在使用梯度下降法训练网络时，实现起来比较简单，同时还有收敛比较快的优点。所以在 sine 函数的拟合中使用 tanh 函数应该是合适的，之后的实践中也证明了这一点，较之前两种训练方式，性能有了极大地提升。

在确定了采用 TanhBPNN 之后，对网络进行训练，通过一次训练之后的验证得到的结果，继续调整参数，训练。直到得到一个理想的结果。（参数的调整过程见 ./log/sine_log.txt 文件）

4.字母识别

1）BP

开始时，使用了 SigmoidBPNN，将一个输入对应的期望输出表现为一个由 0、1 组成的向量，调整了几次后发现效果一般。后来考虑到使用 tanh 函数作为激活函数收敛的更快，因此改为采用了 TanhBPNN，对应的也将期望输出重新表现成了一个-1、1 组成的向量。性能有了一定的提升。

在对输入数据的处理上，为了能够使得数据落在一个比较好的建模区域内，开始时将灰

度值除以 255,使得在（0,1）之间，但是想到这样输入的值都集中在（0,1）上，所以将输入值做处理，为（input-0.5）*2，这样使输入的值在（-1,1）之间。在查阅了书籍后，了解到在对网络的优化上，有一个方法，将激活函数调整为 f(x)=1.7159*tanh(2/3*x)（由 LeCun 提出），原因是这样调整过后使得在（-1,1）间近似为线性增长，使得网络快速收敛。但实践发现性能变化不大。进一步研究发现该方法最好配合着对输入的数据进行一定的处理后使用（消除均值、去相关性、协方差均衡），但限于数学水平有限，无法做到上述的处理，故最终还是采用了 TanhBPNN 的实现。（具体的参数调节件 ./log/letter_log.txt 文件，将调试过程中产生的各个版本的参数配置文件在 ./log/letter_config 中）

2）CNN

这一部分的操作比较少，主要是利用 keras 进行的网络的搭建，原理和 CNN 的图片识别一致。开始时，像官网上的示例一样，采用了“rule”函数作为激活函数，但效果很差，识别率很低，相较于 sigmoid 函数，tanh 函数收敛的更快。共用了三个卷积层，两次池化。池化采用的是最大池化法。

5.分析：

对于 sine 拟合过程中，前面提到的先将期望输出经过变换，再训练最后再变换回来的方法，但结果不理想。我认为主要有两个原因（不见得是对的）：

1）如前文提到，这种变换将原本是一个分布的比较好的数据集合变换到了一个比原来差的分布方式上，导致收敛的比较差。

2）这种变换可以看成是在原本的输出层上又加了一层，该层做了一个线性变换。虽然这层的权重和 bias 不用做调整，但该层会对前面两层的权重的调整产生影响，因为前一层的调整要依赖后一层的参数。如果实现一个有两个隐藏层的 BP 网，按照上述的方式去调整我觉得也能得到一个比较不错的结果。