### 可视化极其重要，没有可视化各个参数的调参属于瞎掰

1.可视化是关键。不要怕浪费时间去写一些好用的训练过程中的可视化工具。如果你还是从terminal中打印出来的loss裸眼的做可视化，那你该考虑一下升级了。

## 模型的持久化，

1. 使用tensorflow中saver=tf.train.Saver()函数创建保存器，使用saver.save()函数可以保存图模型和变量取值，方便在断电或者内存不够时恢复模型。避免因为意外原因没有得到任何结果（任何一次训练都是宝贵的）。

## 使用合适的评价函数来评价模型，

<https://www.coursera.org/learn/deep-neural-network>  [Andrew Ng](https://www.coursera.org/instructor/andrewng)

1. 对于一个模型来说，首先是训练数据和测试数据其次就是评价指标，如果没有明确的评价指标，就像一个无头苍蝇，不知道下一步要调节的是什么，模型在训练集和测试集上出了什么错误。

2.1回归模型可以使用：

R2系数决定系数：总体平方和SStot,残差平方和SSerr R2=1-(SSerrr/SStot)

表示模型对于数据的拟合度，此值越大越好，最好的情况是此值为1.

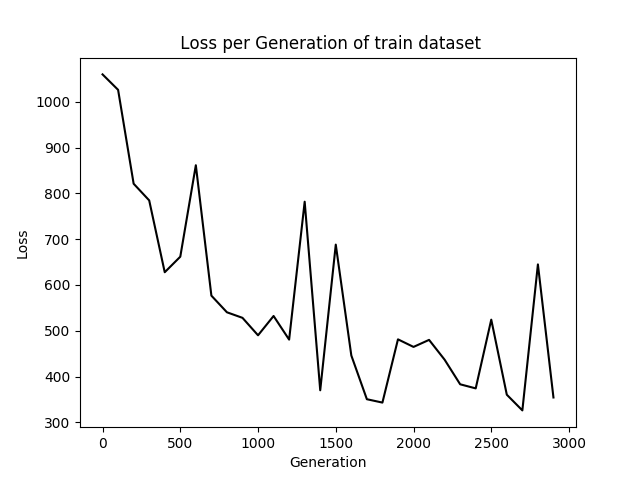
表示标签上的平方差 表示标签和预测值的方差和

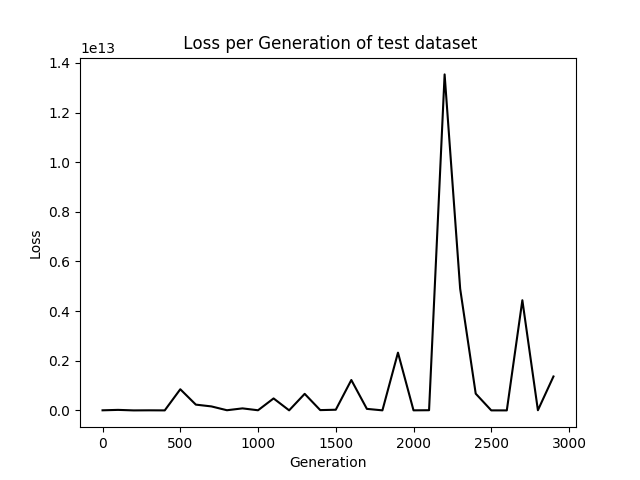
2.2 相对分析误差RPD=(stdev/remse)此值只在测试集上测定，stdev表示测定值标准差，rmse表示测试集上均方根误差。RPD值越大越好。

2.3对于多元素的预测问题，使用多个神经元表示输出，要在不同的元素上分别在训练集和测试集上使用评价指标（因为数据的分布和数据的范围是不同的，但是对于多任务学习其loss值是多种任务loss值的加和，我们不能知道导致loss值大的原因是来自于模型对哪一种预测元素的偏差，只有在多种任务上分别使用测试指标，才能判出模型对于不同任务的准确程度。）

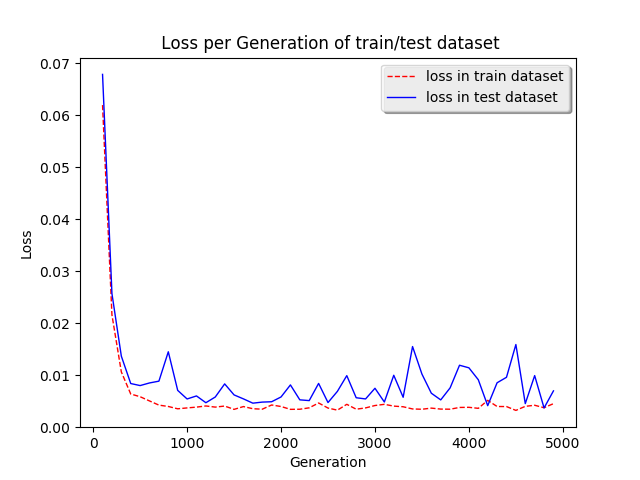
## 数据的预处理，

数据的不同维度一般来说很可能来自不同的数量级，比如目标屈服强度数据范围为200-600，伸长率数据范围为20-60，而硼元素的数据范围一般在0.0005-0.0025之间，如果使用原始数据进行预测则会出现模型不拟合的后果，此时就需要上使用数据归一化处理（max-min归一化，0归一化等等）将数据各维度缩放到一个数据范围内，此处使用的是[0,1]

****

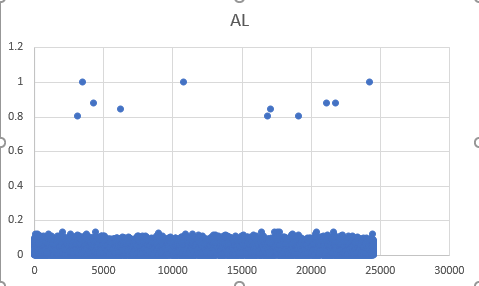
****

以上为不使用归一化的后果。（注意数量级）

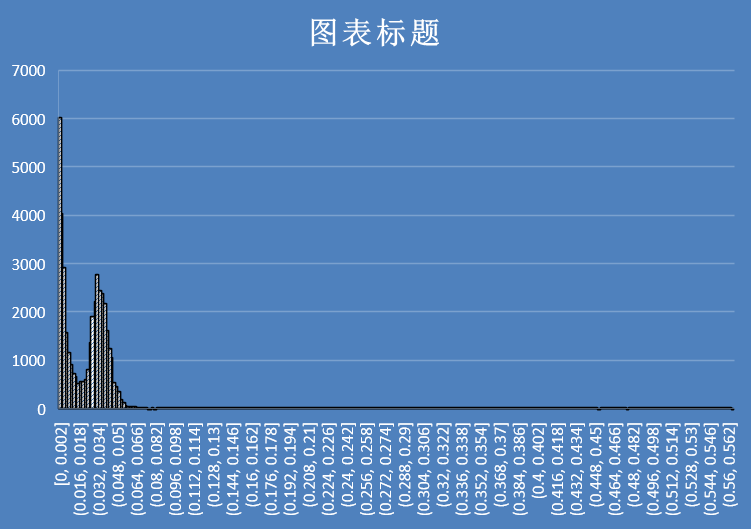


上图是使用数据归一化后的结果

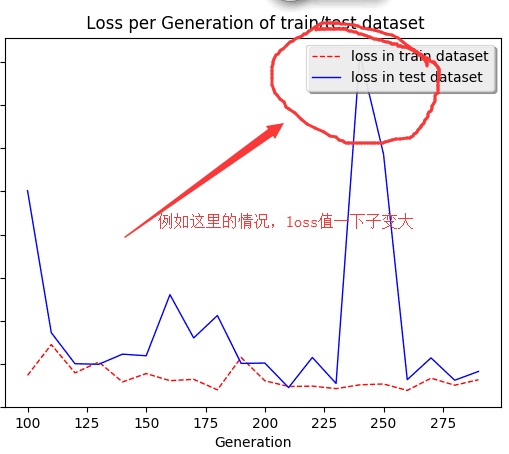
1. 数据一定要保证干净，数据在处理之前要清洗，去除错误的和整体分布明显不一致的数据。（可以使用excel查看csv文件中数据的总体分布）



在归一化的数据后我们发现AL的值一般是分布在0-0.066之间，25000个数据中出现3个值为1的值肯定是数据采集时本身有错误。



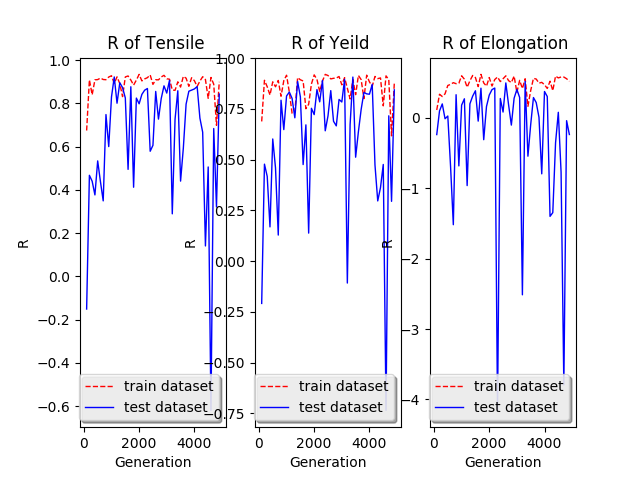
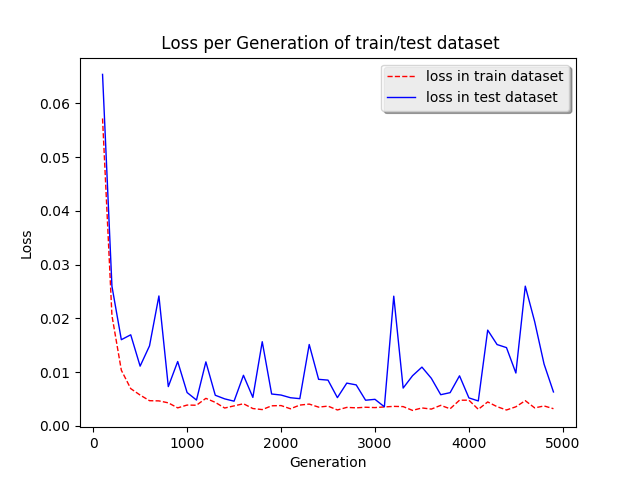
如果这种数据出现在测试集或者验证集中而训练集中没有这种数据很可能出现



Loss值突然一下爆炸的情况

## 可视化与过拟合

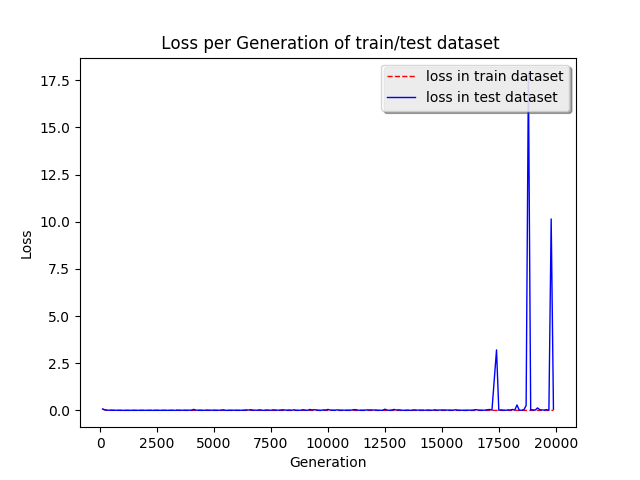
单纯的使用每100或1000次迭代后输出难以看清评价函数的本质特征（毕竟人肉眼观察是十分有限的）。我们应该使用图形的方式展示训练指标函数的变化。

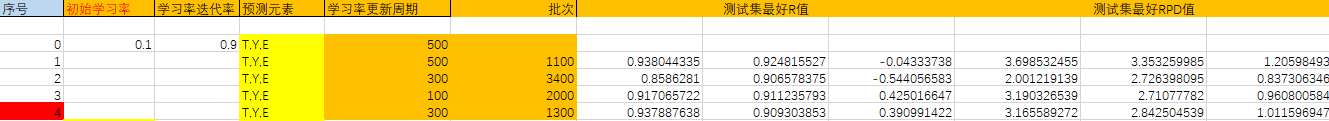


1：可以通过判断训练集和测试集上的误差来判断是否过拟合，或者欠拟合。如果训练集上已经表现不好，则欠拟合，如果训练集表现不错但是测试集上表现太差就过拟合（一般使用L2正则化，Dropout来避免，BN算法也有一定的正则化作用.

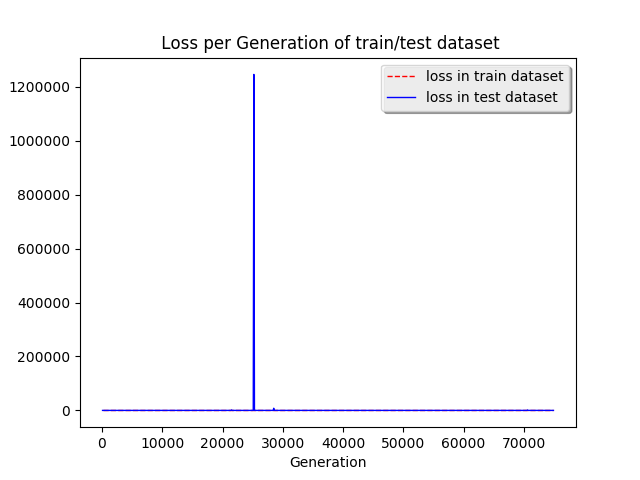
<https://blog.csdn.net/u013555719/article/details/78295927>

2.：方便提前终止训练，因为一般不太复杂的问题，不用迭代很多次，一个不太复杂的模型迭代训练长时间会发生过拟合，因为它会越来越向训练集趋近（即时此时我们加入正则化但是适当的正则化参数是合理的，通常为0.00001-0.0001，如果继续加大一个不太复杂的模型的正则化参数很容易导致模型欠拟合。防止过拟合适可而止，不要因为测试集表现有差距就无限制的加大正则化力度，很容易导致欠拟合）





1. 画图不要等到迭代完成后在调用画图函数，应该在程序运**行时每隔一定时间（每3000或5000）**调用画图函数
2. 避免程序到后期出现bug或者断电等意外而失去了实验结果（每一次训练都是宝贵的）
3. 避免程序到后期输出爆炸而失去原来结果的对比度。（因为数据的数量级相差太大了，原来微小的差异被覆盖掉了）
4. 在合适的时候early stoping，防止过拟合。

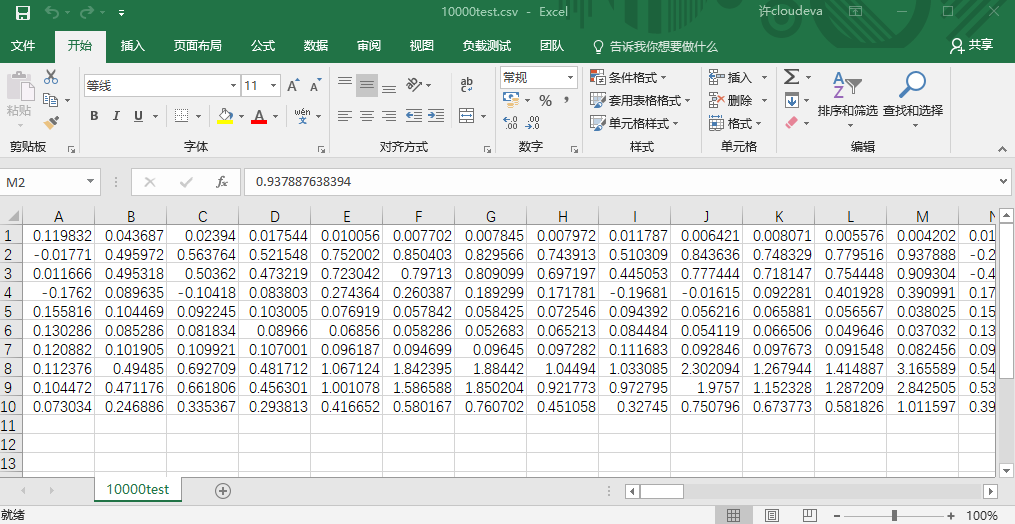


PS:

1.多使用malplotlib观察模型中参数的变化，熟悉excel中各种操作，观察各项指标变化

2.尽量将模型在训练集上表现和测试集上表现呈现在一张图上。

## 训练得到的结果要实时保存（最好5000次迭代保存一次）



例如：上面使我们某次迭代后保存的数据，每一行的值都是一个测试指标而列值则是迭代次数，通过比较分析，我们能够很方便的确认模型的迭代情况。（例如我们经过观察，发现模型在第M列中的2,3行的值是最大的，这表示模型在1300次迭代时有了最好的结果）

## 紧跟开源框架的步伐

Tensorflow是一个开源的框架，其所有代码在github上都有开源，其新版本都有很棒的功能，例如在1.0版本以前，其并没有使用tf.layers.batch\_normalization的函数，使用batch\_normalization需要使用低阶函数tf.nn.batch\_normalization自行定义高级功能。随着其版本的迭代，其功能越来越完善，使用方法越来越便捷，这就需要我们不断的学习与跟进其API。多使用slim,tflearn,tf.layers,keras等高级模块，其对各个函数有良好的封装，尽量少使用或不使用tf.nn函数等低端接口，对于初学者而言很容易写出bug.

## 一定要使用Batch normalization

参考文献：

<https://blog.csdn.net/u013555719/article/details/78388269>

Batch normalization 对于神经网络相当于一个万金油的存在，其能够对每层神经层的输出进行批处理标准化，大大提高训练的速度，精准度，也能为模型带来少量的正则化。注意BN的正确使用方法，BN函数只在训练集时使用，但是在测试集上不使用。测试集上的方差和平均值使用训练集上使用滑动平均算法得到的结果进行估计。

## 调节参数

参考文献：

<http://cs231n.stanford.edu/> 斯坦福大学cs231n李飞飞计算机视觉课程

Deep Learning for Vision: Tricks of the Trade - Bay Area Vision Meeting By Facebook

<https://www.youtube.com/watch?v=clgMTk5V2Sk> 脸书公司2013-2014年机器视觉演讲

A. Krizhevsky, I. Sutskever, and G. E. Hinton. [ImageNet Classification with Deep Convolutional Neural Networks](http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf). In *NIPS*, 2012 Hinton大神的调参心得

[A Brief Overview of Deep Learning](http://yyue.blogspot.com/2015/01/a-brief-overview-of-deep-learning.html/),《神经网络与深度学习》神书 [Ilya Sutskever](http://www.cs.toronto.edu/~ilya/).

K. He, X. Zhang, S. Ren, and J. Sun. [Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification](http://arxiv.org/abs/1502.01852). In *ICCV*, 2015. 何凯明2015论文

<http://lamda.nju.edu.cn/weixs/project/CNNTricks/CNNTricks.html>

南京大学lamda实验室电子书

### 学习率对学习的影响



可以从图中看到低的learn\_rate曲线会趋近于线性（蓝线），其收敛很慢。

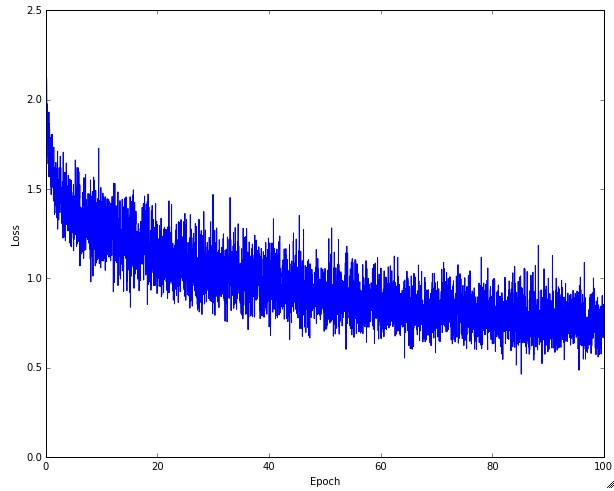
而过高的learn\_rate曲线会趋近于指数（绿线）。会使得曲线迅速下降，但是可能会在一个不好的结果下收敛。过高的学习率甚至会使得结果往差的方向发展（黄线）。

比较合理的学习率应该体现出入红色线所示的形状。

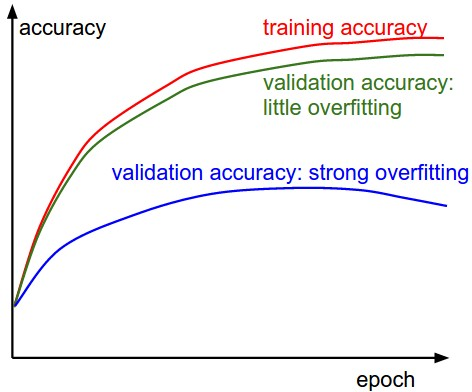
**实际训练中：**

**应该从大的学习速率0.1开始，除以2直到损失没有发散但是注意，如果你更改了学习率参数就不要更改mini-batch参数，因为学习率受mini-batch-size的影响较大**

### Batch size值对学习的影响



实际中学习曲线可能不是平滑的，而是如图上所示有很多的噪声，曲线的震幅反映了batch\_size选择的合理程度，一般batch\_size大的话曲线会变得更加平滑，也就是说训练的参数和实际的参数拟合度更高。  
极端情况是batch\_size的大小和dataset的大小一样，那么曲线很可能就是平滑的。



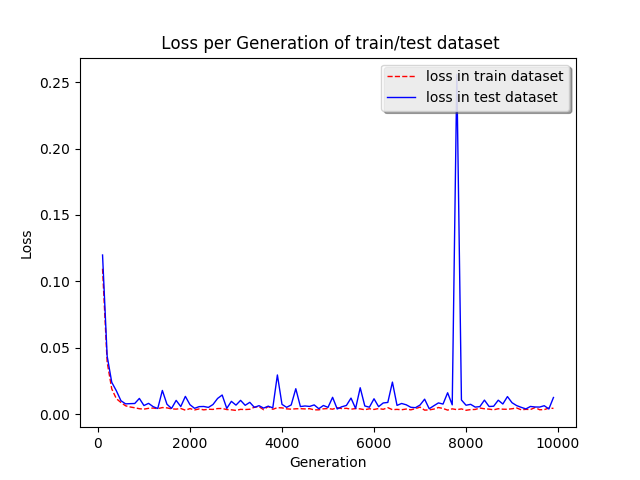
蓝色的曲线显示出验证准确率和训练准确率差别较大，说明可能出现过拟合现象，这种情况可以通过增加正则化项的惩罚系数，或者增加训练样本大小。  
绿色曲线显示出验证准确率和训练准确率差别较小，这虽然是件好事，但是也表明模型的可进步性很小，如果你想尝试取得更好的效果，可以进一步复杂化模型。

**。**

### 更加专注于模型在测试集上的表现

是不是参数越多，或者神经网络模型越深或者网络越宽在测试数据集上表现越好呢？

答案其实是否定的，神经网络模型越深越宽，模型对于训练集上表现会更好，这并不意味着在测试集上也会有相同的表现。比如如果问题本身不是一个很复杂的问题，例如我们的21个维度参数的钢铁性能预测，在我看来和百万人脸识别来说就是一个很小的数据集，我们不用设置很复杂的神经网络去拟合它，因为很容易过拟合。我们5-10层卷积加上5-10层全连接网络在我看来就能实现很好的结果，训练批次在5000以内就能取得很好的结果，如果次数到达10000至以上，在训练集上表现并没有极大的提升，在测试集上却有着及其严重的过拟合。





神经网络大小的选择其实和你需要实现的问题是有关的，并不是说越复杂的神经网络其预测能力就会很好，也有可能会很差。

### 神经网络结构对结果的影响

参考文献

Liu M, Shi J, Li Z, et al. Towards Better Analysis of Deep Convolutional Neural Networks[J]. IEEE Transactions on Visualization & Computer Graphics, 2017, 23(1):91-100.

为了分析卷积神经网络的结构对最后结果的影响，设计了另外两种卷积神经网络来分析网络的深度对结果的影响。表1展示了这三种网络的基本信息。与BaseCNN相比，ShallowCNN少了三层卷积层和一层池化层；DeepCNN的卷积层和池化层数量是BaseCNN的两倍。

网络深度。进一步研究网络的深度对神经元学习特征的影响。

表1不同深度之间的CNN性能比较

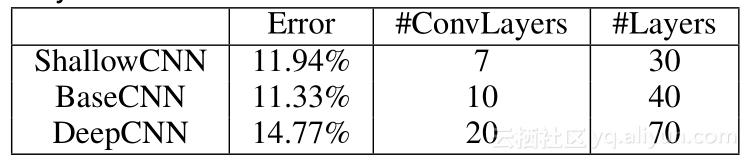
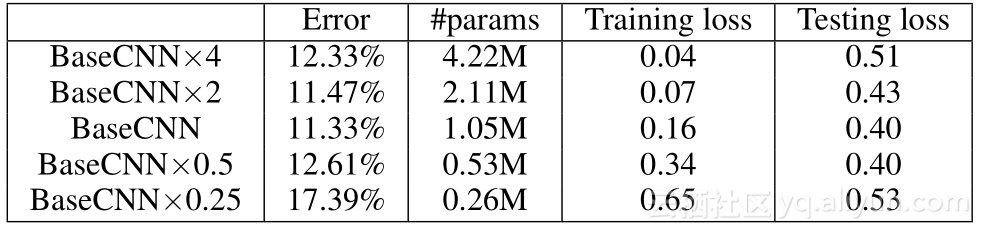


图12 模型深度对性能的影响

发现**如果数据集本身较小，解决的问题不是那么复杂，深层的网络很容易造成过拟合。**在这个例子中，浅层卷积网络和深层卷积网络表现基本相同，更深的网络往往意味着过拟合。

不仅是网络深度对预测结果存在着影响，网络的宽度（即本层神经元中神经元的个数）也有类似的关系。

表2 用于分析卷积神经网络的宽度对结果的影响，BaseCNN#w 表示其每个神经层内神经元的个数是BaseCNN的w倍



从表2可以发现，BaseCNN\*4的训练错误率很低，但是测验错误率却比较高。这说明，这个网络出现了过拟合现象。

### 卷积核和池化层的大小设置

训练期间，输入图像的大小优选为2的幂，诸如32（例如，*CIFAR-10*），*64,224*（例如，常用的*ImageNet*），384或512等。更重要的是，我们应该选用小的卷积核（3\*3）和小的步长（1）以及设置0填充（O padding）这不仅减少了参数的数量，而且提高了整个深度网络的准确率。同时，上面提到的一个特例，即3\*3步长为1的滤镜可以保留图像/特征图的空间大小。对于池化层，常用的池化大小是2\*2。