神经网络之拟合与修正

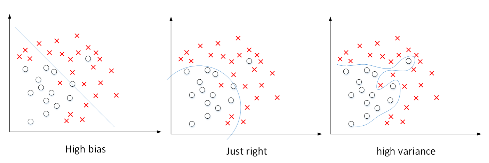
http://blog.csdn.net/red\_stone1/article/details/78208851

1w以上样本，98%训练集，1%交叉验证集（dev），1%测试集。三者务必处于同一分布

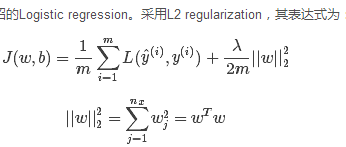
深度学习中的偏差（Bias）和方差（Variance）是抽象意义上的，并不实际指某一数据集分布状态。

偏差指训练集误差与贝叶斯误差（人类最好表现）的差值。偏差大表明欠拟合。解决方法是增大训练次数，增加神经网络隐藏层数、神经元数，或者选择其它更复杂的NN模型

方差指交叉验证集（或测试集）与偏差的差值。方差大表明过拟合。解决方法是增大样本量，使用正则化（Regularization）方法或者dropout方法

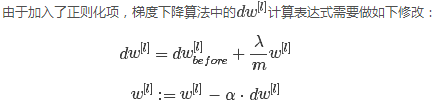


传统机器学习算法中，减小Bias会增加Variance，减小Variance会增加Bias。而现在的深度学习中，通过使用更复杂的神经网络和海量的训练样本，一般能够同时有效减小Bias和Variance。



b值对整体模型影响很小，所以忽略对其正则化

L2 regularization比L1 regularization更加常用，因为其得到的w更加稀疏（很多w为零值），可以节约存储空间；L2在微分求导方面更简单。而且二者解决过拟合问题能力接近。



L2 regularization也被称做weight decay。因为在更新w[l]的时候，会多减去增加的正则项，，使得w[l]比没有正则项的值要小一些

Regulation和dropout方法的本质，都是使训练出的权重参数更小。原因可以从极端情况来思考（这似乎也是神经网络问题常见的思考模式）。当某些权重参数很小，近似为0时，意味着该神经网络中某些神经元的实际作用很小，该神经元可以忽略。从而使得原本过于复杂的神经网络变得简单。

欠拟合与过拟合的本质。借助空间几何理解，欠拟合是采用了过于低维的曲面（比如平面，折面）来划分样本空间中复杂分布的情况。而过拟合则是采用了过于高维的曲面来划分样本空间中不那么复杂的分布情况。维度高低对应着神经网络的复杂程度，直观上就是神经元个数。而神经元个数又是通过权重指数表征的。因此解决过拟合问题，着眼于减小权重参数大小。