1. 验证集和测试集应该\_同一\_\_分布？ （同一/不同）
2. 神经网络模型高方差或高偏差时，分别有哪些可尝试的策略？

Bias反映的是模型在样本上的输出与真实值之间的误差，即模型预测值与真实值的差异，也即算法本身的拟合能力，是由使用的学习算法的某些错误或过于简单的假设造成的误差。它**会导致模型欠拟合，很难有高的预测准确率**。

解决方法：选择新网络，增加隐藏层数量，寻找更具代表性的特征、.用更多的特征（增大输入向量的维度）

Variance反映的是模型每一次输出结果与模型输出期望之间的误差，即不同训练数据训练的模型的预测值之间的差异，也即模型的稳定性，反应预测的波动情况，是由于使用的算法模型过于复杂，导致对训练数据的变化十分敏感，这样会**导致模型过拟合**，使得模型带入了过多的噪音。

解决方法：增大数据集合（使用更多的数据，数据增强）、减少数据特征（减小数据维度）。数据集无法增加则采用正则化或者dropout，或者Early Stopping

1. 梯度消失或爆炸产生的原因及解决方法？

因为网络层数太深而引发的梯度反向传播（链式求导）中的连乘效应。Sigmoid函数图像，呈现一个驼峰状（很像高斯函数），从求导结果可以看出，Sigmoid导数的取值范围在0~0.25之间，而我们初始化的网络权值通常都小于1，因此，当层数增多时，小于0的值不断相乘，最后就导致梯度消失的情况出现。同理，梯度爆炸的问题也就很明显了，就是当权值过大时，导致 ，最后大于1的值不断相乘，就会产生梯度爆炸。

解决办法：

换用Relu，让激活函数的导数为1，就不存在梯度消失爆炸的问题了，每层的网络都可以得到相同的更新速度；

权重正则化（weithts regularization）比较常见的是l1正则，和l2正则；

**梯度剪切**这个方案主要是针对梯度爆炸提出的，其思想是设置一个梯度剪切阈值，然后更新梯度的时候，如果梯度超过这个阈值，那么就将其强制限制在这个范围之内。这可以防止梯度爆炸。