处理过拟合的方法——权重衰退

权重衰退是常见的处理过拟合的一种方法。是常用的正则化方法之一，通常也被称为L2正则化

正则化是处理过拟合的常用方法：在训练集的损失函数中加入惩罚项，以降低学习到的模型的复杂度。

保持模型简单的一个特别的选择是使用L2惩罚的权重衰减。这会导致学习算法更新步骤中的权重衰减。

权重衰减功能在深度学习框架的优化器中提供。

在同一训练代码实现中，不同的参数集可以有不同的更新行为。

暂退法（Dropout）丢弃法

丢弃法是在层之间加入噪音，本质上是一个正则，将一些输出项随机置0来控制模型复杂度

通常将丢弃法作用在隐藏全连接层上，很少应用在CNN上

推理中的丢弃法：正则项只在训练中使用，只会对权重产生影响，在预测的时候权重不需要发生变化的情况下不需要正则，所以推理过程不需要正则，确保在推理过程中能够有一个确定性的输出

数值稳定性：

神经网络的梯度：网络层数较多时向量每次进行求导计算会产生新的矩阵，根据链式法则，网络中将会产生大量的矩阵相乘计算导致梯度爆炸和梯度消失。这是数值稳定性的两个常见问题。

优化训练方法可以防止这两个问题，例如将乘法变成加法

此外可以通过类似生成随机变量使每一层的元素值都相似，主要用均值和方差这两个特点控制相似度，同时，对反向传播的梯度值也进行相似的归一化。

权重初始化：

在合理值区间里随机初始参数。合理的权重初始值和激活函数的选取可以提升数值稳定性。

1. 权重衰退：方法介绍及代码的简单实现
2. 丢弃法：作用层、推理中的丢弃法及丢弃法的代码实现
3. 数值稳定性：神经网络梯度、优化方法、权重初始化