

# 深度学习面试常见问题

## bias/variance trade off ?

## ReLU 在零点不可导，那么在反向传播中如何处理？

可以在 ReLU 的零点人为的给它赋值一个倒数，比如 0 或者 1

## ReLU 的优缺点

所有的负值都变为 0，而正值不变，这种操作被称为**单侧抑制**。正因为有了单侧抑制，才使得神经网络中的神经元也有了稀疏激活性。

使用 ReLU 得到的 SGD 的收敛速度会比 sigmoid/tanh 快很多，计算复杂度低，不需要进行指数运算。

训练的时候很脆弱，很容易“die”。同时 ReLU 不会对数据做幅度压缩，所以数据的幅度会随着模型层数的增加不断扩张

## 四种归一化

- BN, batch normalization
- LN, layer normalization
- IN, Instance normalization, 实例归一化
- GN, group normalization

## 什么是梯度消失和梯度爆炸

关键在于激活函数。在反向传播中，我们使用的是梯度下降策略来优化参数，在链式法则求导的时候，涉及到了对激活函数求导，关键在于反向传播时激活函数的导数。如果导数大于 1，那么随着网络层数的增加，梯度更新将会朝着指数爆炸的方式增加，这就是梯度爆炸。同样如果导数小于 1，那么对着网络层数的增加梯度更新信息会朝着指数衰减的方式减少，这就是梯度消失。

## 什么是梯度弥散，，造成的问题

梯度弥散就是梯度消失，导数为0。梯度弥散造成靠近输出层的隐藏层梯度大，参数更新快，所以很快就会收敛。靠近输入层的隐藏层梯度小，参数更新慢，几乎和初始状态一样，随机分布。

## 如何解决梯度消失问题

- 使用 ReLU, Leaky ReLU 激活函数替代 Sigmoid
- BN, 通过每一层的输出规范为均值和方差一致的方法, 消除了权重参数放大缩小带来的影响

## 什么是端到端学习 (end-to-end)

端到端学习是一种解决问题的思路, 与之对应的是多步骤解决问题, 依旧是将一个问题拆分为多个步骤分布解决, 而端到端是由输入端的数据直接得到输出端的结果。

## Softmax 的原理是什么, 有什么作用

softmax 用于多分类过程中, 它将多个神经元的输出, 映射到  $(0, 1)$  区间内, 可以看成概率来理解, 从而进行多分类。

softmax 直白来讲就是将原来输出是 3, 1, -3 通过 softmax 函数作用, 就映射为  $(0, 1)$  的值, 而且这些值类和为 1 (满足概率的性质), 那么我们就可以将它理解为概率, 在最后选取输出节点的时候, 我们就可以选取概率最大 (也就是值对应最大) 的节点, 作为我们的预测目标。

softmax 的好处:

1. 好求导
2. 使得好结果和坏结果之间的差异更加显著, 更有利于学习。SVM 只选自己喜欢的男神, softmax 把多有备胎全都拉出来评分, 最后还归一化一下。

## dropout 是否用在测试集上?

不能, 只能用于训练集, 是训练过程中的一种正则化技术

## 列举几个梯度下降的方法并简介

## 常见降维方法, 并简介