

问题

1、什么叫做稀疏权重矩阵

在矩阵中，若数值为0的元素数目远远多于非0元素的数目，并且非0元素分布没有规律时，则称该矩阵为稀疏矩阵；与之相反，若非0元素数目占大多数时，则称该矩阵为稠密矩阵。定义非零元素的总数比上矩阵所有元素的总数为矩阵的稠密度。

2、relu函数为什么实现了去线性化

3、如何理解梯度消失和梯度爆炸

4、如何理解L1、和L2正则化

5、为什么L1正则化有助于生成一个稀疏权值矩阵

L1正则化有助于生成一个稀疏权值矩阵，进而可以用于特征选择。为什么要生成一个稀疏矩阵？

稀疏矩阵指的是很多元素为0，只有少数元素是非零值的矩阵，即得到的线性回归模型的大部分系数都是0。通常机器学习中特征数量很多，例如文本处理时，如果将一个词组（term）作为一个特征，那么特征数量会达到上万个（bigram）。在预测或分类时，那么多特征显然难以选择，但是如果代入这些特征得到的模型是一个稀疏模型，表示只有少数特征对这个模型有贡献，绝大部分特征是没有贡献的，或者贡献微小（因为它们前面的系数是0或者是很小的值，即使去掉对模型也没有什么影响），此时我们就可以只关注系数是非零值的特征。这就是稀疏模型与特征选择的关系。

6、为什么参数值小的模型比较简单

拟合过程中通常都倾向于让权值尽可能小，最后构造一个所有参数都比较小的模型。因为一般认为参数值小的模型比较简单，能适应不同的数据集，也在一定程度上避免了过拟合现象。可以设想一下对于一个线性回归方程，若参数很大，那么只要数据偏移一点点，就会对结果造成很大的影响；但如果参数足够小，数据偏移得多一点也不会对结果造成什么影响，专业一点的说法是『抗扰动能力强』。

7、假设神经网络有很多层，对a3这个隐藏层进行反向随机失活之后，a3的期望值为什么不变

8、为什么drought使得代价函数不再明确定义

9、什么是对抗代价。

10、为什么L2也被称为权重衰减法，它很像梯度下降？