

# 神经网络参数初始化方法

原创

十三吖

于 2022-02-07 15:10:08 发布

1009

★ 收藏 1

版权

分类专栏:

机器学习相关

算法工程师面试

推荐算法与Tensorflow

文章标签:

神经网络

深度学习

网络参数

网络权重

参数初始化



机器学习相关 同时被 3 个专栏收录

5 订阅

92 篇文章

订阅专栏

**神经网络** 训练的过程就是对网络权重不断学习更新的过程，网络初始权重对网络的训练非常重要。不合适的初始化方法可能会导致网络参数传播过程中产生梯度消失、梯度爆炸等现象。

常用的初始化方法有随机初始化、Xavier初始化、he初始化等

## 1 零初始化

对于**逻辑回归**，网络权重是可以初始化为0的；对于深度神经网络，网络权重和偏置是不可以一起初始化为0的，不然会造成每层的网络所有节点输出是一致的，具体分析可以参考[神经网络权重为什么不能初始化为0?](#)。

## 2 随机初始化

随机初始化的时候常常采用高斯或**均匀分布**初始化网络权重。这种方法相对0初始化要好许多，但是在遇到激活函数为sigmoid / tanh的时候，可能会出现**梯度消失和梯度爆炸现象**。

以四层网络，参数为 $w_1, b_1, w_2, b_2, w_3, b_3, w_4, b_4$ ，激活函数为sigmoid， $\sigma(x) = \frac{1}{1+e^{-x}}$ 。

$$y_i = \sigma(z_i)$$

$$z_i = w_i y_{i-1} + b_i$$

sigmoid函数求导数 $\sigma' = \sigma(1 - \sigma)$ 后，峰值为0.25；损失函数C对 $b_1$ 的导数为

$$\frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial y_4} \frac{\partial y_4}{\partial z_4} \frac{\partial z_4}{\partial y_3} \frac{\partial y_3}{\partial z_3} \frac{\partial z_3}{\partial y_2} \frac{\partial y_2}{\partial z_2} \frac{\partial z_2}{\partial y_1} \frac{\partial y_1}{\partial z_1} \frac{\partial z_1}{\partial b_1} = \frac{\partial C}{\partial y_4} \sigma'(z_4) w_4 \sigma'(z_3) w_3 \sigma'(z_2) w_2 \sigma'(z_1) 1$$

sigmoid 函数对于大到10的值，sigmoid的值几乎是1，对于小到-10的值，sigmoid的值几乎为0。意味着如果权值矩阵被初始化成过大的值，权重 $w_i$ 连乘会出现梯度爆炸的现象，反之，当权值矩阵被初始化成太小的值，可能会出现梯度消失

## 3 Xavier初始化

Xavier初始化通过保持输入和输出的方差一致（服从相同的分布）避免梯度消失和梯度爆炸问题，使得信号在神经网络中可以传递得更深，在经过小或太大）。



十三吖

关注

xavier均匀分布

$$w \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_{in} + n_{out}}}, \frac{\sqrt{6}}{\sqrt{n_{in} + n_{out}}}\right]$$

xavier正态分布

$$w \sim N\left[\text{mean} = 0, \text{std} = \frac{\sqrt{2}}{\sqrt{n_{in} + n_{out}}}\right]$$

适用于激活函数为tanh的深层网络，但不适用于RELU

## 4 He初始化

He初始化解决的问题：ReLU网络每一层有一半的神经元被激活，另一半为0（x负半轴中是不激活的），所以要保持variance不变，只需要在Xavier的基础上再除以2。

He均匀分布

$$w \sim U\left[-\frac{\sqrt{6}}{2\sqrt{n_{in} + n_{out}}}, \frac{\sqrt{6}}{2\sqrt{n_{in} + n_{out}}}\right]$$

He正态分布

$$w \sim N\left[\text{mean} = 0, \text{std} = \frac{\sqrt{2}}{\sqrt{n_{in}}}\right]$$

## 参考

- 1、[参数初始化](#)
- 2、[网络权重初始化方法总结](#)

欢迎关注微信公众号（算法工程师面试那些事儿），本公众号聚焦于算法工程师面试，期待和大家一起刷leetcode，刷机器学习、深度学习面试题等，共勉～



十三吖

关注



关注

## 神经网络参数初始化总结分析

weixin\_42521239的博客 · 2993

神经网络的参数主要是权重 (weights) :  $W$ , 和偏置项 (bias) :  $b$ 。训练神经网络的时候需先给定一个初...

## 【深度学习】神经网络参数的初始化方法

小马日记 · 2291

一、全0初始化 全0初始化最为简单, 但不适合采用。因为如果将神经网络中参数全部初始化为0, 那么每个...

## 神经网络 参数初始化方法

qq\_39809262的博客 · 224

参数初始化方法 文章目录参数初始化方法1. 使用NumPy来初始化2. [Xavier 初始化方法](http://proceedings...

## 神经网络参数初始化 (MSRA、Xavier)

qq\_37705280的博客 · 389

再推荐一篇感觉原理讲的很好的文章: <https://blog.csdn.net/u014696921/article/details/53819512>

## 5.3 神经网络参数初始化 最新发布

dc12499574的博客 · 277

文章目录一、内置随机化 init函数二、内置常数化 init函数三、不同层 or 块进行不同的init四、自定义init函数...

## 神经网络中的参数的初始化

weixin\_45268911的博客 · 590

神经网络中的参数初始化 任何网络在训练之前, 都需要经过参数初始化, 在神经网络中, 权重初始化方法对...

## 神经网络初始化

qq\_42527487的博客 · 232

神经网络参数初始化非常重要, 适合的初始化可以简化训练过程, 提高精确率; 参数初始化不是一成不变的...

## 神经网络参数初始化方式

七七的博客 · 4226

看了文章《Understanding the difficulty of training deep feedforward neural networks》, 里面提出了两种参数...

## 神经网络的初始化方法总结 | 又名“如何选择合适的初始化方法”

CV技术指南 (微信公众号) · 349

前言本文介绍了为什么初始化很重要, 总结了常用的几种初始化方法: 全零或等值初始化、正态初始化、均...

## 机器学习笔记 (9) — 神经网络的参数初始化、架构选择、训练过程

weixin\_45718019的博客 · 348

本文主要介绍神经网络的参数初始化、架构选择、训练过程 1.神经网络的参数初始化 我们以下网络为例: 在...

## 常用神经网络参数初始化方法

倾城之恋的专栏 · 295

一般神经网络被随机初始化。随机从均匀分布或者高斯分布中采样。 1、Xavier初始化 适用于tanh, sigmoid...

## 神经网络的参数初始化

zhuguqin1的专栏 · 1725

注: 本文是学习吴恩达老师机器学习公开课的学习笔记。1. 对于神经网络fminunc函数, 我们需要初始化神...

## sigmoid函数\_温故知新——激活函数及其各自的优缺点

weixin\_39945445的博客 · 2208

1.什么是激活函数? 所谓激活函数 (Activation Function) , 就是在人工神经网络的神经元上运行的函数, 负...

## 神经网络参数初始化-He初始化

u010505915的博客 · 4052

Glorot和Bengio提出了Xavier方法来初始化网络参数。该方法基于激活是线性的假设。但该假设对ReLU不适...

“相关推荐”对你有帮助么?



非常没帮助



没帮助



一般



有帮助



非常有帮助

©2022 CSDN 皮肤主题: 编程工作室 设计师: CSDN官方博客 返回首页

关于我们 招贤纳士 商务合作 寻求报道 400-660-0108 kefu@csdn.net 在线客服 工作时间 8:30-22:00

公安备案号11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 经营性网站备案信息

北京互联网违法和不良信息举报中心 家长监护 网络110报警服务 中国互联网举报中心 Chrome商店下载

©1999-2022北京创新乐知网络技术有限公司 版权与免责声明 版权申诉 出版物许可证 营业执照



十三吖  
码龄6年

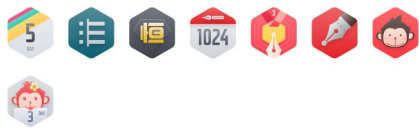
新手认证



十三吖

关注

原创 周排名 总排名 访问 等级  
7893 997 440 194 2138  
积分 粉丝 获赞 评论 收藏



私信

关注

搜博主文章



### 热门文章

推荐算法常用评价指标：NDCG、MAP、MRR、HR、ILS、ROC、AUC、F1等 55004

LDA主题模型及python实现 33615

oracle（数据库创建、删除、启动、关闭） 22220

python使用turtle库与random库绘制雪花 21650

时间序列的距离度量DTW 17340

### 分类专栏

	机器学习相关	92篇
	算法工程师面试	41篇
	推荐算法与Tensorflow	104篇
	Leetcode	2篇
	诗和远方	2篇
	PyQt5	19篇



十三吖



十三吖

关注

weixin\_49094019: 为什么我运行有好多问题啊 🤔

中医证型关联规则挖掘

#404: 请问 inputfile = './data/apriori.txt' #输入事务集文件这个文件是怎么生成的?

LDA主题模型及python实现

Zaxhoil: 博主, 您好, 我想请问一下您, 这些输出的数字的含义是什么, 怎么理解?

知识图谱与推荐系统之《Multi-Task Feat...

meirucio: www会议

基于用户的协同过滤算法

就是个狠人: 你写得很棒

您愿意向朋友推荐“博客详情页”吗?



强烈不推荐



不推荐



一般般



推荐



强烈推荐

## 最新文章

推荐系统中常用的embedding方法

对比学习的一个原则3个关键点

transformer - Attention is all you need

2022年 12篇

2021年 40篇

2020年 2篇

2019年 57篇

2018年 157篇

2017年 2篇

## 目录

1 零初始化

2 随机初始化

3 Xavier初始化

4 He初始化

参考



十三吖

关注