

全连接神经网络剖析(2)-重复

该课程主要为大家讲授如下的内容：

- 全连接神经网络的基本概念
- 激活函数
- 全连接神经网络的参数
- 全连接神经网络的前向传播过程
- 神经网络的优化问题

1. 全连接神经网络的基本概念

前馈神经网络（feedforward neural network）是一种非常简单神经网络，参数从输入到输出逐层单向传递。

各神经元分层排列

每个神经元只与前一层的神经元相连

- 神经元接收前一层的输出，并输出给下一层
- 各层间没有反馈
- 内部不会构成有向环

前馈神经网络的结构

- 输入层（1层）
- 隐层（0~* 层）
- 输出层（1层）

(c)Microsoft 2022. 仅供个人学习使用

多层感知机（Multilayer Perceptron, MLP）是一种前馈神经网络。除了输入层，它每层的节点（神经元）都带有非线性的激活函数。

- 容易混淆的概念
 - “多层感知机” \neq 多层的单感知机
 - 单感知器
 - 使用阈值激活函数
 - 执行二进制分类
 - 多层感知器
 - 可以使用任意激活函数
 - 激活函数可以自由地执行分类或者回归任务

MLP

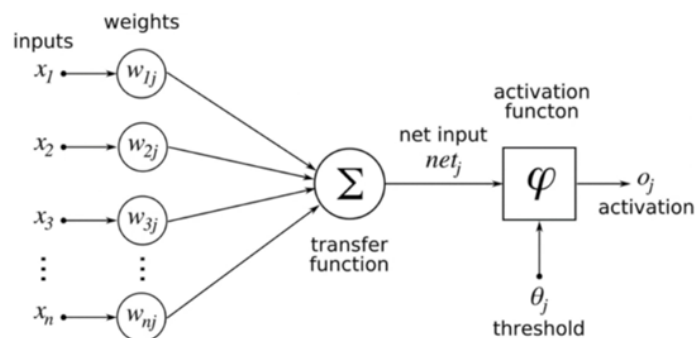
多层

前馈神经网络

全连接神经网络是一种多层感知机。全连接神经网络中，除了输入层，每一层的任意节点都和上一层的每一个节点相连。每个节点存储一个标量，这个标量是同上一层所有节点连接的权值分别与各个节点的输出值相乘后的线性组合结果。

全连接神经网络的神经元

- 每个神经元
 - 将前一层的输入线性加权求和
 - 通过非线性激活函数输出



2. 激活函数

全连接神经网络是一种多层感知机。全连接神经网络中，除了输入层，每一层的任意节点都和上一层的每一个节点相连。每个节点存储一个标量，这个标量是同上一层所有节点连接的权值分别与各个节点的输出值相乘后的线性组合结果。

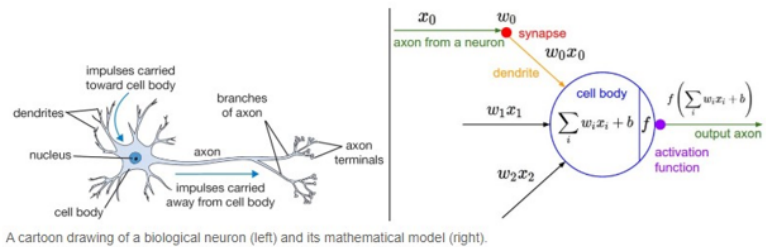
为什么必须引入激活函数

没有激活函数

- 全连接神经网络的每一层仅作线性运算
- 逐层叠加的结果：线性运算

激活函数

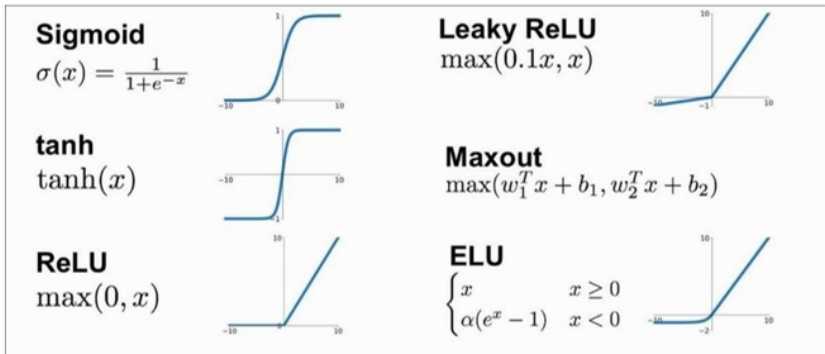
- 引入非线性因素
- 拟合能力更强



(c)Microsoft 2022. 仅供个人学习使用

常用的激活函数有Sigmoid函数、双曲正切函数、ReLU函数等。

常见激活函数



(c)Microsoft 2022. 仅供个人学习使用

3. 全连接神经网络的参数

全连接神经网络的参数是其中每一个神经元的参数的集合。每一个神经元都有以下参数：

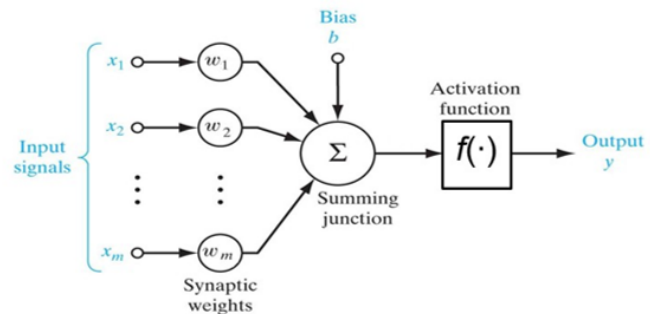
1. 权重 (weights)

2. 偏置 (bias)

全连接神经网络的参数

Microsoft AI Talent Program

- 每个神经元的参数
 - 权重: w_1, w_2, \dots, w_m
 - m 个权重
 - m = 前一层神经元数量
 - 偏置: b
 - 1 个偏置
- FC NN的参数
 - 所有神经元的权重和偏置



(c)Microsoft 2022. 仅供个人学习使用

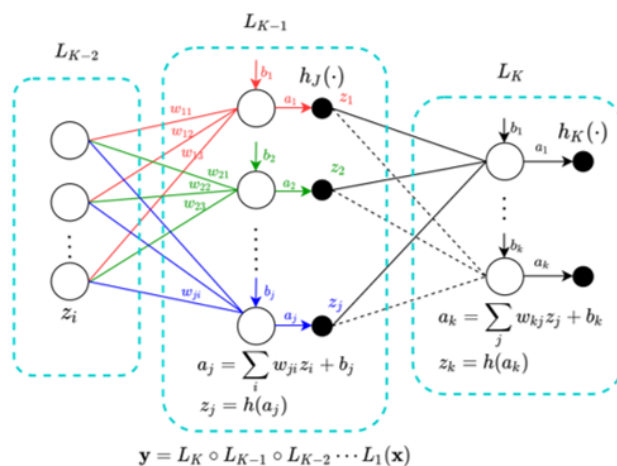
4. 神经网络的前向传播过程

神经网络作为统计机器学习模型，它的推理过程就是前向传播计算的过程。全连接神经网络从输入层接受参数以后，每一层都的每一个神经元都完成接受上一层输入-计算神经元值-激活函数变换-输出的过程，直到输出层。

全连接神经网络的使用

Microsoft AI Talent Program

- 结构
- 激活函数
- 参数值



(c)Microsoft 2022. 仅供个人学习使用

5. 神经网络的优化问题

神经网络模型的优化过程就是所谓的“训练”过程。优化的目的就是要令最终输出与期望值之间的距离最小，评估这个距离的是损失函数。由于神经网络参数多，更新量大，一般采用反向传播算法进行训练。

神经网络训练过程

Microsoft AI Talent Program



算出 \hat{y}

损失函数算出 L

反向传播
更新参数

反复迭代

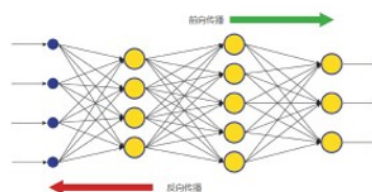
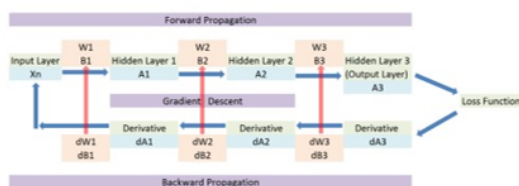
(c)Microsoft 2022. 仅供个人学习使用

反向传播算法（Back Propagation, BP）是梯度下降算法在多层神经网络上的应用。它根据损失函数得到的误差来反向逐层传播梯度并更新参数。

反向传播

Microsoft AI Talent Program

- 通过前向传播求出损失函数
- 以最小化 损失函数 为目的，逐层向前
- 每层对参数求梯度，按照减小损失函数的方向更新权值和偏置



(c)Microsoft 2022. 仅供个人学习使用

神经网络模型的优化问题可能不是一个凸优化问题，因此它的“训练结

果”并不一定是全局最优解。

神经网络的损失函数

Microsoft AI Talent Program

- 非凸函数
- 非凸优化 vs 凸优化

