

# 第四次作业

徐薪-519021910726

提交 DDL: 2021 年 12 月 30 日 0 时

作业完成形式有三种:

- (1) 你可以手写自己的解答并拍照, 再将照片整理成一份 word/pdf 文件并提交。
- (2) 你可以使用 word 文档进行编辑, 最后提交 word/pdf 文件。
- (3) 你可以使用 latex 进行编辑, 最后提交 pdf 文件。

如果你没有在 DDL 之前提交作业, 请及时在微信群里联系助教进行补交。如果对作业有任何问题, 你可以在从微信里询问助教谢瑜璋, 或者发邮件到 [constantjxyz@sjtu.edu.cn](mailto:constantjxyz@sjtu.edu.cn)。

## 1 本次作业可能用到的知识点

本次作业可能会用到以下知识点:

- (1) 机器学习的基础知识, 过拟合、欠拟合的定义, bias-variance 平衡。
- (2) 神经网络的结构, 神经网络反向传播的推导。
- (3) 层次化聚类、kmeans 聚类、dbscan 聚类的定义、方法、优缺点。

## 2 第一题

解释机器学习模型的过拟合 (overfitting) 与欠拟合 (underfitting) 指的是什么现象。我们可以有哪些方法来避免过拟合与欠拟合的现象。

**Solution.** 过拟合和欠拟合是模型泛化能力低的两种原因。过拟合是指模型过度训练数据集本身, 从而忽略了它的规律, 把一些不好的甚至是不应该的特性都学习进来了, 从而会出现模型在训练集上的误差远远小于测试集上的误差。欠拟合是指模型拟合程度不高, 数据距离拟合曲线较远, 或指模型没有很好地捕捉到数据特征, 不能够很好地拟合数据。在欠拟合的情况下, 模型在训练集上的表现差, 在测试集上表现同样会很差。增加数据量可以避免欠拟合问题, 正则化方法、dropout 可以避免过拟合问题。合理的数据切分可以避免欠拟合和过拟合的问题。□

## 3 第二题

根据以下的步骤完成神经网络反向传播的推导。假定一个前馈全连接神经网络的结果如下图所示, 其中,  $(x_{k-1,1}, x_{k-1,2}, \dots, x_{k-1,N_{k-1}})$  为该神经网络第  $k-1$  层共  $N-1$  个神经元的输出信号, 并被输入第  $k$  层神经元。对于第  $k$  层的第  $j$  个神经元, 根据神经网络的前向信号传播规律, 我们规定

$$net_{k,j} = \left[ \sum_{i=1}^{N_{k-1}} (m_{k,j,i} \cdot x_{k-1,i}^3 + n_{k,j,i} \cdot x_{k-1,i}) \right] + b_{k,j} \quad (1)$$

(1) 假定第  $k$  层的激活函数是 sigmoid 函数, 那么第  $k$  层第  $j$  个神经元输出的信号  $x_{k,j}$  等于什么? (可以用  $net_{k,j}$  表示结果)。

**Solution.**  $x_{k,j} = \frac{1}{1+e^{-net_{k,j}}}$  □

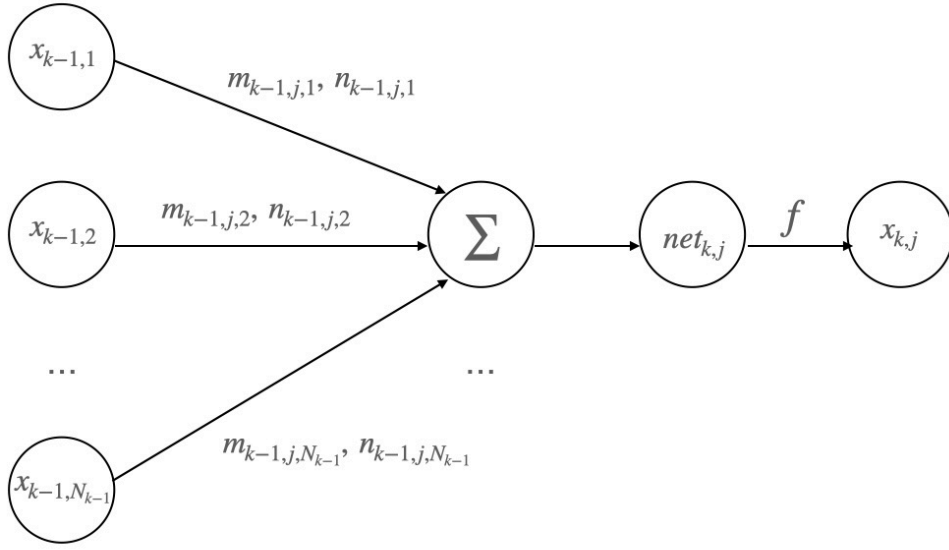


图 1: 第二题的神经网络

(2) 假定我们计算该网络的输出后, 得到其 loss 为  $E$ , 我们第  $k$  层的第  $j$  个神经元从网络输出处反向传播而来的梯度为

$$\frac{\partial E}{\partial x_{k,j}} = G_{k,j} \quad (2)$$

请计算对于  $net_{k,j}$  的反向梯度  $\frac{\partial E}{\partial net_{k,j}}$ 。要求用  $G_{k,j}$  与  $x_{k,j}$  表示该反向梯度。

提示: sigmoid 函数求导公式:  $\frac{\partial \text{sigmoid}(x)}{\partial x} = \text{sigmoid}(x) \cdot [1 - \text{sigmoid}(x)]$ 。

**Solution.**  $\frac{\partial E}{\partial net_{k,j}} = \frac{\partial E}{\partial x_{k,j}} * \frac{\partial x_{k,j}}{\partial net_{k,j}} = G_{k,j} * x_{k,j} * (1 - x_{k,j})$  □

(3) 利用上方计算好的梯度  $\frac{\partial E}{\partial net_{k,j}}$  表达式计算参数  $u_{k,j,i}$  处的梯度  $\frac{\partial E}{\partial m_{k,j,i}}$ 。要求用  $G_{k,j}$  与  $x_{k,j}$  表示该反向梯度。

**Solution.**  $\frac{\partial E}{\partial m_{k,j,i}} = \frac{\partial E}{\partial net_{k,j}} * \frac{\partial net_{k,j}}{\partial m_{k,j,i}} = \frac{\partial E}{\partial net_{k,j}} * x_{k-1,i}^3 = G_{k,j} * x_{k,j} * (1 - x_{k,j}) * x_{k-1,i}^3$  □

(4) 假如该神经网络的学习率为  $\eta$ , 那么参数  $m_{k,j,i}$  更新后的值  $m'_{k,j,i}$  为多少? 要求用  $G_{k,j}$ ,  $x_{k,j}$ ,  $m_{k,j,i}$  与  $\eta$  表示更新后的参数。

**Solution.**  $m'_{k,j,i} = m_{k,j,i} - \eta * \frac{\partial E}{\partial m_{k,j,i}} = m_{k,j,i} - \eta * (G_{k,j} * x_{k,j} * (1 - x_{k,j}) * x_{k-1,i}^3)$  □

## 4 第三题

我们在 lecture14 中学习了层次化聚类 (hierarchical clustering, HAC), k-means, dbscan 密度聚类三种聚类方法, 请比较一下这三种算法的时间复杂度和优缺点 (假定计算两点之间的距离时间复杂度为  $O(1)$ )。

**Solution.** 1. HAC

时间复杂度:  $O(n^3)$  或  $O(n^2 \log n)$

优点: 无参数; 可解释性强。

缺点: 复杂度高。

2. k-means

时间复杂度:  $O(Knm)$

优点: 复杂度较低; 对凸型分布的数据有较好的分类效果。

缺点: 收敛速度受初始值影响; 对除了凸型分布数据以外的其他类型数据分类效果较差。

3. dbscan

时间复杂度:  $O(n^2)$

优点: 可过滤噪声; 适用于多种类型分布的数据。

缺点: 对初始参数设置敏感。

□

## 5 作业反馈

点击访问链接<https://www.wjx.cn/vj/Q0makrC.aspx>或者扫描下方二维码就可以反馈意见。



图 2: 作业调查问卷