

第四次作业

人工智能 CS410 2021年秋季

姓名：李子龙 学号：518070910095 日期：2021 年 12 月 29 日

题目 1. 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。

解. 过拟合是指训练误差和测试误差之间的差距太大。为了避免过拟合，可以采用正则化方法，提高模型的泛化能力：参数正则化、Early Stop和Dropout、数据集增强等。

欠拟合是指模型不能在训练集上获得足够低的误差。为了避免欠拟合，可以增加网络复杂度或者增加特征。 □

题目 2. 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

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

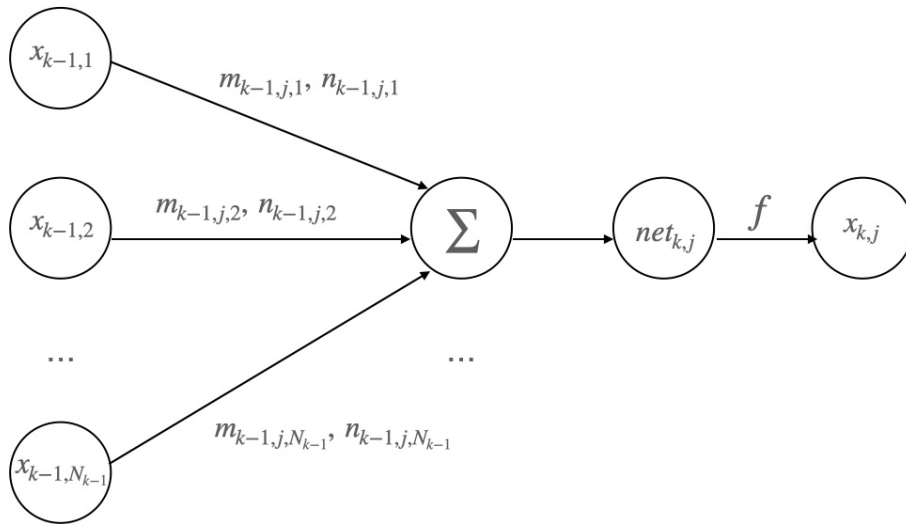


图 1: 第二题的神经网络

- (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）。

- (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： $\frac{\partial \text{sigmoid}(x)}{\partial x} = \text{sigmoid}(x) \cdot [1 - \text{sigmoid}(x)]$ 。

- (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。
- (4) 假如该神经网络的学习率为 η ，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 η 表示更新后的参数。

解. (1)

$$x_{k,j} = \text{sigmoid}(net_{k,j}) = \frac{1}{1 + e^{-net_{k,j}}}$$

(2)

$$\begin{aligned} \frac{\partial E}{\partial net_{k,j}} &= \frac{\partial E}{\partial x_{k,j}} \frac{\partial x_{k,j}}{\partial net_{k,j}} \\ &= G_{k,j} \text{sigmoid}(net_{k,j}) [1 - \text{sigmoid}(net_{k,j})] \\ &= G_{k,j} x_{k,j} (1 - x_{k,j}) \end{aligned}$$

(3)

$$\begin{aligned} \frac{\partial E}{\partial m_{k,j,i}} &= \frac{\partial E}{\partial net_{k,j}} \frac{\partial net_{k,j}}{\partial m_{k,j,i}} \\ &= G_{k,j} x_{k,j} (1 - x_{k,j}) x_{k-1,i}^3 \end{aligned}$$

(4)

$$\begin{aligned} m'_{k,j,i} &= m_{k,j,i} - \eta \frac{\partial E}{\partial m_{k,j,i}} \\ &= m_{k,j,i} - \eta G_{k,j} x_{k,j} (1 - x_{k,j}) x_{k-1,i}^3 \end{aligned}$$

□

题目 3. 我们在lecture14中学习了层次化聚类（hierarchical clustering, HAC）, k-means, db-scan 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 $O(1)$ ）。

解. 如下表所示。

	时间复杂度	优点	缺点
HAC	$O(n^2) \sim O(n^3)$	可解释性好	需要计算每个样本间的距离，复杂度高
k-means	$O(tkmn)$	复杂度小	无法作用于非凸簇
DBSCAN	$O(n \log n) \sim O(n^2)$	聚类可以不局限于球形	需要手动选择参数 ϵ 和 MinPts

□