

# 人工智能基础

## 作业 6

注意：

- 1) 请在网络学堂提交**电子版**；
- 2) 请在**11月24日晚23:59:59**前提交作业，**不接受补交**；
- 3) 5道题目中任选3道解答(多做不加分；5题全做则按题目的解答顺序，只计前3题的分数，如提交作业中题目解答顺序是1、3、2、4, 则第4题不计分)。
- 4) 如有疑问，请联系助教：

杨鹏帅: yps18@mails.tsinghua.edu.cn

鄞启进: yqj17@mails.tsinghua.edu.cn

崔雪建: cuixj19@mails.tsinghua.edu.cn

高子靖: gzj21@mails.tsinghua.edu.cn

鲁永浩: yonghao.lu@foxmail.com

牛家赫: njh20@mails.tsinghua.edu.cn

江 澜: jiangl20@mails.tsinghua.edu.cn

尹小旭: yxx21@mails.tsinghua.edu.cn

1. 使用二元 Logistic 回归公式的对数模型进行扩展，推导 Softmax 回归模型，即下式：

$$P(y^{(i)} = k) = \frac{e^{\beta_k x_i}}{\sum_{k=1}^K e^{\beta_k x_i}}$$

提示：使用线性模型和归一化因子对结果为  $k$  的概率的对数进行建模。

$$\log P(y^{(i)} = k) = \beta_k x_i - \log Z$$

2. 试设计一个前馈神经网络来解决 XOR（异或）问题，要求该前馈神经网络具有两个隐藏神经元和一个输出神经元，要求激活函数为 ReLU。
  - a) 请给出你设计的网络的具体参数，并验证其能够满足上述要求。
  - b) 请证明：若网络的输入限制为两输入( $x_1, x_2$ )，且激活函数为线性函数，则无法解决上述异或问题。

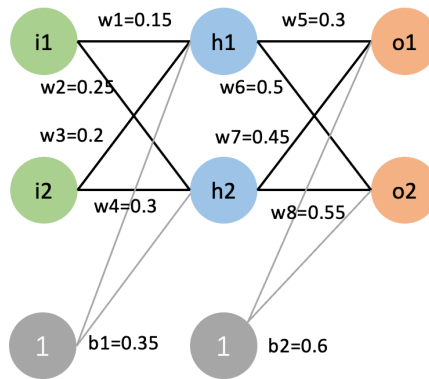
3. 有关神经网络的正则化，请回答下列问题：

- a) 正则化通过降低模型的复杂度，提高泛化能力，请从该角度解释为什么正则化时加入惩罚项的参数不包括偏置  $b$ 。
- b) 当我们使用 L2 范数作为惩罚项时，其目标函数变为

$$C = C_0 + \lambda \sum_{i=1}^n \omega_i^2$$

其中  $C_0$  为原目标函数， $\lambda > 0$  为规范化参数。若采用梯度下降算法，请写出此时的权重更新的表达式，并分析其能够提高模型泛化能力的原因。

- c) 在实际模型训练过程中，如何选取正则化参数  $\lambda$ ，请给出一种思路。
4. 现有两输入的前馈神经网络及其参数如下所示，且隐藏层和输出层神经元均使用了 Sigmoid 激活函数，请回答如下问题：



- a) 当输入为  $(i1, i2) = (0.05, 0.10)$  时，计算该神经网络的输出分别为多少。请写明必要的计算过程。
  - b) 在 a) 的基础上，若  $(o1, o2) = (0.12, 0.95)$ ，采用最小化均方误差作为优化准则，请根据反向传播算法计算参数  $w5$  和  $w6$  的梯度。
  - c) 在 b) 的基础上，若采用梯度下降更新参数，且学习率设为 0.1，写出更新后的参数  $w5$  和  $w6$ 。
5. 关于激活函数，请回答如下的问题：
- a) 若在多层的神经网络中使用  $\text{Tanh}$ （双曲正切单元）作为神经元的激活函数，请根据 BP 算法推导出参数梯度的表示形式，说明其可能导致梯度消失问题的原因，并提出一种解决方案。
  - b) “死亡 ReLU”问题是指当使用 ReLU 作为激活函数时，大梯度流经某个神经元后，容易导致神经元输出始终为零。请举例说明该问题，并给出两种解决思路。
  - c) 对于下述的两种激活函数

$$\text{swish}(x) = x \cdot \text{sigmoid}(\beta x)$$

$$\text{GELU}(x) = xP(X \leq x)$$

其中  $\beta$  为常数， $X$  服从标准高斯分布，其概率密度函数为  $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ 。

请求出这两种激活函数的导数，试比较 GELU 与 ReLU 激活函数，并说明 GELU 激活函数的优势。