1 Homework for the chapter of MLP

problem 4.1

感知机（perceptron）是一种线性分类的模型，即求出能将数据进行线性划分的超平面，用数学公式表示如下：

$$f(x) = sign(w \cdot x + b)$$

这个三维的 XOR 问题可以看做为三维空间中的单位正方体的 8 个角，如下图所示:

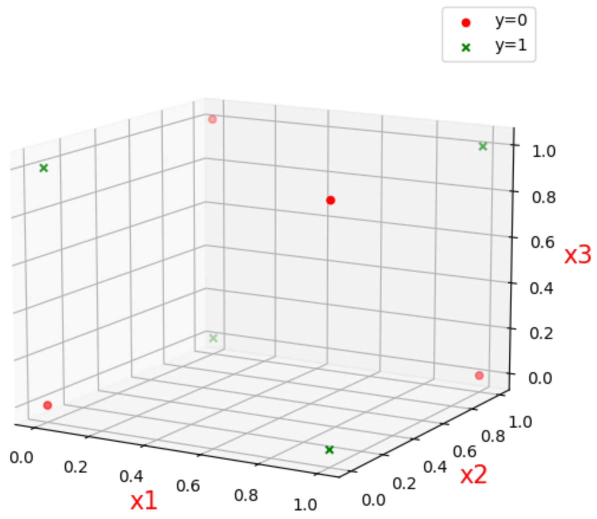

图 1: 三维空间中点的分布

而使用感知机的求解过程就是找到一个决策面能将所有的 $y = 0$ 的点与 $y = 1$ 的点划分开。但是从上图可以看出，无论取什么样的二维平面，都无法将红色和绿色的点完美划分开，所以只用感知机是无法解决该 XOR 问题。

problem 4.2

首先，假设第二层的任意神经元为：

$$f_i^{(2)}(x) = g(w_i^{(2)} \cdot x + b_i^{(2)})$$

其中，$g(\cdot)$ 为线性激活函数，x 为输入层。

则任意第三层的神经元可以表示为:

$$f_j^{(3)}(X) = g'(w_j^{(3)} \begin{bmatrix} f_1^{(2)}(x) \\ f_2^{(2)}(x) \\ ... \\ f_n^{(2)}(x) \end{bmatrix} + b_j^{(3)})$$

上式可以表示为:

$$f_j^{(3)}(X) = g'(w_j^{(3)} \begin{bmatrix} g(w_1^{(2)} \cdot x + b_1^{(2)}) \\ g(w_2^{(2)} \cdot x + b_2^{(2)}) \\ ... \\ g(w_n^{(2)} \cdot x + b_n^{(2)}) \end{bmatrix} + b_j^{(3)})$$

又由于激活函数 $g(\cdot)$ 为线性函数,则 $g(\cdot)$ 可以表示为:

$$g(x) = ax + b$$

则第三层的神经元表示最终化简如下:

$$f_j^{(3)}(X) = g'(w_j^{(3)} \begin{bmatrix} w_1'^{(2)} \\ w_2'^{(2)} \\ ... \\ w_n'^{(2)} \end{bmatrix} \cdot x + w_j^{(3)} \begin{bmatrix} b_1'^{(2)} \\ b_2'^{(2)} \\ ... \\ b_n'^{(2)} \end{bmatrix} + b_j^{(3)})$$

最终形式类似单层:

$$f_j^{(3)}(X) = g'(w_j'^{(3)} \cdot x + b_j'^{(3)})$$

则可以看出使用线性激活函数连接的多层神经元可以视为一层。所以这个三层的多层感知机可以视为两层 (仅包含输入层和输出层),而且根据问题 4.1 的结论,两层的感知机无法解决 XOR 问题和 n 位奇偶校验问题。

problem 4.3

(a) 每个隐藏单元与所有的输入单元连接,输出单元所有隐藏单元连接,则权重数量为:

$$n_{weights} = d \cdot n_H + n_H \cdot c$$

(b) 根据 (a) 中的权重数量结果,发现权重数量与隐藏层神经元数量呈线性相关,其表达式如下:

$$n_{weights} = (d + c) \cdot n_H$$

problem 4.4

sigmoid 函数的导数用其本身表示如下:

$$\begin{aligned}
f'(x) &= (\frac{1}{1+e^{ax}})' = -a\frac{e^{ax}}{(1+e^{ax})^2} = -a\frac{1+e^{ax}-1}{(1+e^{ax})^2} \\
&= af(x)(f(x)-1)
\end{aligned}$$