

机器学习概论第一次作业

刘浚哲

北京大学物理学院 1500011370

March 31, 2018

1. 感知器模型简述¹

感知器 (perceptron) 是单层神经网络当中最基本的工作单元. 假设输入空间 (特征空间) 是 $\chi \in R^n$, 输出空间是 $\gamma = \{+1, -1\}$. 输入 $x \in \chi$ 表示实例的特征向量, 对应输入空间 (特征空间) 的点; 输出 $y \in \gamma$ 表示实例的类别. 由输入空间到输出空间的如下函数:

$$f(x) = \text{Act}(w \cdot x + b) \quad (1)$$

其中, w 为权值, b 为偏置项, $\text{Act}(x)$ 函数称为激活函数, 其可以取不同的函数类型. 感知器的假设空间为: 定义在特征空间中的所有线性分类模型 (linear classification model) 或线性分类器 (linear classifier), 即函数集合 $\{f|f(x) = w \cdot x + b\}$.

从感知器的定义看来, 它是一个线性分类器, 因此其学习策略是根据数据集的线性可分性来实现的: 给定某一训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (2)$$

其中 $x_i \in \chi = R^n, y_i \in \gamma = \{+1, -1\}, i = 1, 2, \dots, n$, 若存在某个超平面 (w^*, b^*) :

$$w^* \cdot x + b^* = 0 \quad (3)$$

能够将数据集的正例点和负例点完全划分到超平面的两侧, 则称数据集 T 为线性可分数据集.

2. 异或问题

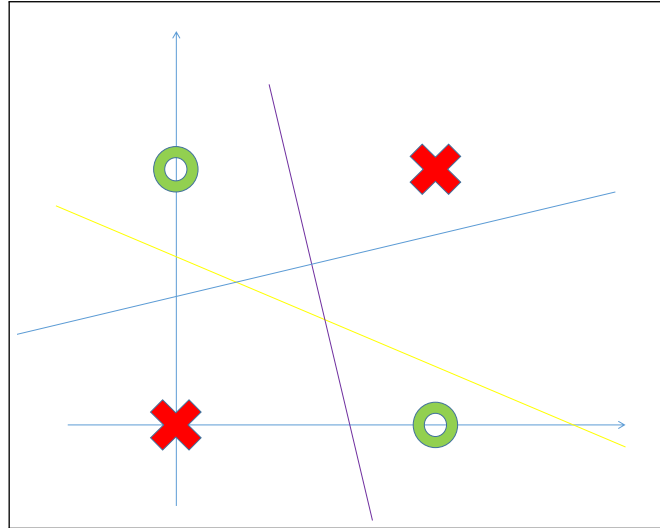
所谓异或问题, 就是两个输入 x_1, x_2 当中有且只有一个输入为 1 时, 输出才为 1, 否则为 0. 列表如下所示:

异或问题 (XOR)

输入 x_1	输入 x_2	输出 y
0	0	0
0	1	1
1	0	1
1	1	0

如果要将一个异或问题的数据集交给感知器 (单层神经网络) 去处理的话, 则会出现下图的结果:

¹ Email Address: 1500011370@pku.edu.cn



感知器解决异或问题

很显然, 这个数据集在二维输入的情形下正例反例是不能被一条直线所分开的, 因此异或问题的数据集是非线性可分的, 因此不能用感知器来处理. 事实上, 只要是单层的神经网络, 或者是线性分类模型, 都无法处理异或分类问题.

简单的数学证明如下: 假设存在 w_1, w_2, b 满足异或分类条件, 则由前三式可以得到约束条件:

$$b < 0 \quad (4)$$

$$w_1 + b > 0 \quad (5)$$

$$w_2 + b > 0 \quad (6)$$

三式合起来则有:

$$w_1 + w_2 + b = (w_1 + b) + (w_2 + b) + (-b) > 0 \quad (7)$$

这与最后一式:

$$w_1 + w_2 + b < 0 \quad (8)$$

相矛盾. 故不存在这样的模型参数能够实现异或问题.

3. 感知器实现异或问题

单层神经网络无法实现异或问题, 但能够实现线性可分的问题: 例如且 (AND), 或 (OR), 非 (NOT) 等问题. 而我们通过逻辑学知识知道, 异或问题可以拆为:

$$x_1(XOR)x_2 = (x_1(OR)x_2)(AND)((NOT\ x_1)(OR)(NOT\ x_2)) \quad (9)$$

由此, 我们可以多加一层隐层, 来实现或操作和非或操作, 再在输出层实现且操作, 即可实现异或问题.