

Chapter 2 : Perceptron

图灵2201 张祎迪

2.1 Minsky 和 Papert 指出:感知机因为是线性模型，所以不能表示复杂的函数，如异或(XOR)。验证感知机为什么不能表示异或。

对异或操作，训练数据集 $T = \{((0,0),0),((1,1),0),((0,1),1),((1,0),1)\}$

用反证法，如果感知机可以表示异或，那么 $\exists w = (w_1, w_2)^T \in \mathbf{R}^2, b \in \mathbf{R} \text{ st.}$

$$\begin{cases} w \cdot (0,0)^T + b < 0 \\ w \cdot (1,1)^T + b < 0 \\ w \cdot (0,1)^T + b > 0 \\ w \cdot (1,0)^T + b < 0 \end{cases} \quad (1)$$

该方程组无解，因此对于任意的 $w \in \mathbf{R}^2, b \in \mathbf{R}$ ，感知机模型 $f(x) = \text{sign}(w \cdot x + b)$ 都存在误分类点，即感知机不能表示异或。

2.3 证明以下定理:样本集线性可分的充分必要条件是正实例点集所构成的凸壳与负实例点集所构成的凸壳互不相交。

设集合 $S \subset \mathbf{R}^n$ 是由 \mathbf{R}^n 中 k 个点所组成的集合，定义 S 的凸壳为：

$$\text{conv}(S) = \{x = \sum_{i=1}^k \lambda_i x_i \mid \sum_{i=1}^k \lambda_i = 1, \lambda_i \geq 0, i = 1, 2, \dots, k\} \quad (2)$$

充分性：

$$\text{dist}(x_i, x_j) = \|x_i - x_j\|_2$$

定义 $\text{conv}(S_+)$ 和 $\text{conv}(S_-)$ 之间的距离为：

$$\text{dist}(\text{conv}(S_+), \text{conv}(S_-)) = \min \|s_+, s_-\| (s_+ \in \text{conv}(S_+), s_- \in \text{conv}(S_-)) \quad (1)$$

设 $x_+ \in \text{conv}(S_+), x_- \in \text{conv}(S_-)$ 满足 $\text{dist}(x_+, x_-) = \text{dist}(\text{conv}(S_+), \text{conv}(S_-))$

$$\therefore \text{conv}(S_+) \cap \text{conv}(S_-) = \emptyset$$

$$\therefore \text{dist}(\text{conv}(S_+), \text{conv}(S_-)) > 0$$

$$\therefore \forall x \in \text{conv}(S_+), \text{dist}(x, x_-) > \text{dist}(x, x_+)$$

$$\forall x \in \text{conv}(S_-), \text{dist}(x, x_+) > \text{dist}(x, x_-)$$

$$\therefore \exists w = x_+ - x_-, b = -\frac{x_+ \cdot x_+ - x_- \cdot x_-}{2} \text{ st}$$

$$\begin{aligned} \forall x \in \text{conv}(S_+), w^T x + b &= (x_+ - x_-) \cdot x - \frac{x_+ \cdot x_+ - x_- \cdot x_-}{2} \\ &= \frac{\|x - x_-\|^2 - \|x - x_+\|^2}{2} > 0 \end{aligned}$$

$$\begin{aligned} \forall x \in \text{conv}(S_-), w^T x + b &= (x_+ - x_-) \cdot x - \frac{x_+ \cdot x_+ - x_- \cdot x_-}{2} \\ &= \frac{\|x - x_-\|^2 - \|x - x_+\|^2}{2} < 0 \end{aligned}$$

$\therefore S_+, S_-$ 线性可分

必要性:

若 S_+, S_- 线性可分, 则 $\exists w^T x + b = 0$ 将 S_+, S_- 分开, 则 $\forall x_i \in S_+, w^T x_i + b = \varepsilon_i > 0, i = 1, 2, \dots, p$

$\therefore \forall s_+ \in \text{conv}(S_+)$:

$$\begin{aligned} w^T s_+ &= w^T \sum_{i=1}^p \lambda_i x_i \\ &= \sum_{i=1}^p \lambda_i (\varepsilon_i - b) \\ &= \sum_{i=1}^p \lambda_i \varepsilon_i - b \sum_{i=1}^p \lambda_i \\ &= \sum_{i=1}^p \lambda_i \varepsilon_i - b \end{aligned}$$

$\therefore w^T s_+ + b > 0$ 类似地 $w^T s_- + b < 0$ for $\forall s_- \in \text{conv}(S_-)$

若 $\text{conv}(S_+) \cap \text{conv}(S_-) \neq \emptyset$, 则 $\exists s \in \text{conv}(S_+) \cap \text{conv}(S_-)$ st $w^T s + b > 0$ and $w^T s + b < 0$

推出矛盾, 所以可得 $\text{conv}(S_+) \cap \text{conv}(S_-) = \emptyset$