

感知机

W.J.Z

摘要：感知机基于误分类损失函数，利用梯度下降法极小化损失函数，求得将特征空间进行线性划分的超平面，具有原始形式和对偶形式，是神经网络与支持向量机的基础。

$\|w\|$ 为 w 的 L_2 范数 (就是点到平面的距离公式)。

对于误分类实例 (x_i, y_i) ，则公式 4 成立，该点到超平面的距离为公式 5：

$$-y_i (w \cdot x + b) > 0 \quad (4)$$

$$-\frac{1}{\|w\|} y_i (w \cdot x_i + b) \quad (5)$$

1 感知机

感知机公式：

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

w 为权重向量， x 为特征向量， b 为偏置。sign 函数为公式 2：

$$\text{sign}(x) \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (2)$$

从公式 1 可以看出，感知机所做的工作就是求得一超平面，即获取 w 和 b 的值，将所有实例划分为两类，位于超平面上方的为正类，位于超平面下方的为负类。

对于所有的误分类点集合 M ，则 M 到超平面的总距离为公式 6：

$$-\frac{1}{\|w\|} \sum_{x_i \in M} y_i (w \cdot x_i + b) \quad (6)$$

不考虑 $\frac{1}{\|w\|}$ ，则得到感知机的损失函数公式 7。

$$L(w, b) = - \sum_{x_i \in M} y_i (w \cdot x_i + b) \quad (7)$$

损失函数在正确分类时为 0，给定训练数据集，损失函数 $L(w, b)$ 是 w, b 的连续可导函数。

损失函数的梯度由公式 8 和公式 9 所示：

$$\nabla_w L(w, b) = - \sum_{x_i \in M} y_i x_i \quad (8)$$

$$\nabla_b L(w, b) = - \sum_{x_i \in M} y_i \quad (9)$$

2 学习策略

给定一个数据集：

$$T = \{(x_1, y_1) \cdots (x_n, y_n)\}$$

空间上任意一点 x_i 到超平面的距离公式 3 所示：

$$\frac{1}{\|w\|} (w \cdot x_i + b) \quad (3)$$

w, b 更新策略：

$$w \leftarrow w + \eta y_i x_i$$

$$b \leftarrow b + \eta y_i$$

η ($0 < \eta \leq 1$) 成为学习率。

2.1 感知机原始形式算法

Algorithm 1: 感知机原始形式学习算法

Input: 训练数据集 T , 学习率 η

Output: $w, b, f(x) = \text{sign}(w \cdot x + b)$

- 1 初始化 w_0, b_0
- 2 在训练集中选取 (x_i, y_i)
- 3 如果 $y_i(w \cdot x_i + b) \leq 0$

$$w \leftarrow w + \eta y_i x_i$$

$$b \leftarrow b + \eta y_i$$

- 4 返回第 2 步, 直到完全分类正确。
-

2.3 感知机对偶形式算法

Algorithm 2: 感知机对偶形式学习算法

Input: 训练数据集 T , 学习率 η

Output: $\alpha, b,$

$$f(x) = \text{sign}(\sum_{j=1}^N \alpha_j y_j x_j \cdot x + b)$$

- 1 $\alpha \leftarrow 0, b \leftarrow 0$
- 2 在训练集中选取 (x_i, y_i)
- 3 如果 $y_i(\sum_{j=1}^N \alpha_j y_j x_j \cdot x + b) \leq 0$

$$\alpha_i \leftarrow \alpha_i + \eta$$

$$b \leftarrow b + \eta y_i$$

- 4 返回第 2 步, 直到完全分类正确。
-

2.2 感知机对偶模式

什么是对偶模式: 将 w, b 表示为 (x_i, y_j) 的线性组合的形式, 通过求解系数来达到求解 w, b 的目的, 如公式 10 和公式 11 所示。

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (10)$$

$$b = \sum_{i=1}^N \alpha_i y_i \quad (11)$$

$\alpha_i \geq 0$. 从结果输出来看, 原始模式和对偶模式一个输出 w, b , 一个输出 α, b , 二者差别在于求 w 的方法不同, 不同的初始值和不同选取误分类点的次序都会造成最后的解不尽相同。

2.4 算法收敛性

感知机算法收敛性证明请查看李航的统计学习算法。水平有限, 总结如有错误之处请谅解。