



分享人: zyj 2019/6/20





- 任务:二分类
- 模型:线性分类模型 $f(x) = w \cdot x + b$
- 策略:基于误分类的损失函数
- 算法:梯度下降法极小化损失函数
- 感知机对应于输入空间中将实例划分为正负两类的分离超平面,属于判别模型
- 是神经网络和支持向量机的基础





- 輸入空间: χ⊆ Rⁿ 输出空间: {-1,+1}
- 由输入空间到输出空间的函数:

•
$$f(x) = sign(w \cdot x + b)$$

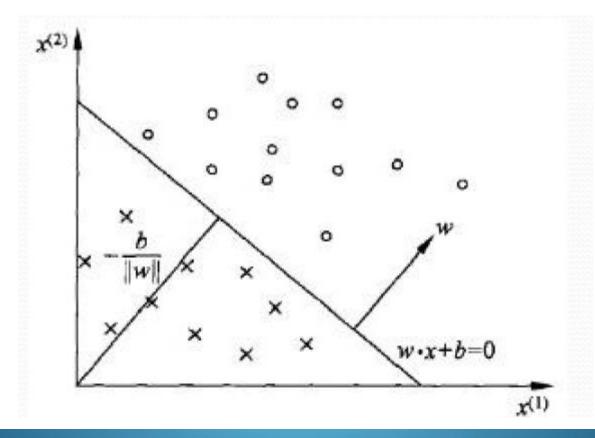
- 称为感知机。
- 模型的参数即权值向量 w 和偏置 b

• 符号函数
$$sign(x) = \begin{cases} +1, & x \ge 0 \\ -1, & x < 0 \end{cases}$$





- 学习得到 R^n 输入空间中一个 R^{n-1} 的超平面。
- · w 为超平面的法向量, b 为超平面的截距





- 即如何定义损失函数。
- 误分类点的数目。

$$Loss = \sum_{i=1}^{n} L_{0-1}(f(x_i), y_i)$$

缺陷?



- 误分类点到超平面的总距离。
- · 样本点 x_i 到超平面的距离:

$$d(x_i; w, b) = \frac{1}{\| w \|} |w \cdot x_i + b|$$



- 求点 x_i 到超平面 $w \cdot x + b = 0$ 的距离 d
- 设点 x_i 在超平面上的投影为 x_0 ,则 $w \cdot x_0 + b = 0$
- $|w\overline{x_0x_i}| = |w| \cdot |\overline{x_0x_i}| = ||w|| \cdot d$
- $|w\overline{x_0}\overline{x_i}| = |w \cdot (\overline{x_i} \overline{x_0})| = |w \cdot x_i w \cdot x_0|$ $= |w \cdot x_i (-b)| = |w \cdot x_i + b|$

$$||w|| \cdot d = |w \cdot x_i + b|$$

$$d = \frac{1}{||w||} |w \cdot x_i + b|$$





$$d = \frac{1}{\|w\|} |w \cdot x_i + b|$$

• 对于误分类的点

$$-y_i(w \cdot x_i + b) > 0$$

则有

$$d(x_i; w, b) = -\frac{1}{\|w\|} y_i(w \cdot x_i + b)$$

因此 误分类点到超平面总距离 = $-\frac{1}{\|w\|}\sum_{x_i\in M}y_i(wx_i+b)$

Loss = $-\sum_{x_i \in M} y_i(wx_i + b)$



• $Loss = -\sum_{x_i \in M} y_i(wx_i + b)$

• 如何优化求得 w 和 b,使得 Loss 最小化

$$\min_{w,b} - \sum_{x_i \in M} y_i(wx_i + b)$$



• 随机梯度下降法

- 随机初始化 w 和 b,即任意选择一个超平面 $w \cdot x + b = 0$
- 对于新的误分类样本 (x_i, y_i) , 计算 w 与 b 的梯度:

$$\frac{\partial L}{\partial w} = -y_i x_i$$

$$\frac{\partial L}{\partial h} = -y_i$$

• 沿负梯度方向更新 w 与 b

$$w_i \leftarrow w_{i-1} + \eta y_i x_i$$

$$b_i \leftarrow b_{i-1} + \eta y_i$$

• 重复以上两个步骤直至样本点被全部正确分类





• 算法的收敛性(即模型在有限次迭代后收敛),证明见书。

- 结论:
- 对于线性可分数据集,感知机算法在有限次迭代后收敛。
- 模型的解不唯一,取决于初始值以及每次更新的样本选择。

- 如何唯一确定分类的超平面?
- · 对于线性不可分数据集,模型 Loss 的表现?





- 基本思想:将 w 与 b 看作是样本的线性组合,因此模型的参数变为了样本的组合系数。
 - 由

$$w_i \leftarrow w_{i-1} + \eta y_i x_i$$
$$b_i \leftarrow b_{i-1} + \eta y_i$$

可得

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad b = \sum_{i=1}^{N} \alpha_i y_i$$

其中, $\alpha_i = a_i \eta$, a_i 可以看作是样本 x_i 被误分类的次数

• 参数更新规则:

$$\alpha_i \leftarrow \alpha_{i-1} + \eta \quad b \leftarrow b_{i-1} + y_i$$

• 代入w和b的公式,得感知机

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

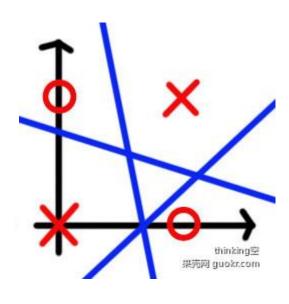
Gram 矩阵
$$G = [x_i \cdot x_j]_{N \times N}$$

对偶形式相比原始形式,大大提升了计算效率。





- 局限





- 任务:二分类
- 假设:数据集线性可分
- 假设空间: 所有线性分类器 $\{f(x)|f(x) = w \cdot x + b\}$
- 策略:

$$Loss = -\frac{1}{\|w\|} \sum_{x_i \in M} y_i (wx_i + b)$$

• 算法:梯度下降法



