

支持向量机(SVM)

1. 线性可分支持向量机

- 支持向量机最简单的情况就是线性可分支持向量机，或硬间隔支持向量机，条件是训练数据线性可分，学习策略是最大间隔法，可表示为凸二次规划问题。

- 原始最优化问题为

$$\begin{aligned} \min_a \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i(w * x_i + b) - 1 \geq 0 \end{aligned}$$

- 求得的最优化问题的解为 w^*, b^* ，得到线性可分支持向量机，分离超平面为

$$w^* * x + b^* = 0$$

分类决策函数为

$$f(x) = \text{sign}(w^* * x + b^*)$$

- 线性可分支持向量机的最优解存在且唯一，位于间隔边界上的实例点为支持向量，最优分离超平面由支持向量完全决定。
- 二次规划的对偶问题是

$$\begin{aligned} \min_a \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j (x_i * x_j) - \sum_{i=1}^N a_i \\ \text{s.t.} \quad & \sum_{i=1}^N a_i y_i = 0 \\ & a_i \geq 0, i = 1, 2, \dots, N \end{aligned}$$

- 通过求解对偶问题学习线性可分支持向量机，首先求得最优值 a^* ，然后求最优值 w^*, b^* ，得出分离超平面和分类决策函数。

2. 线性支持向量机

- 训练数据近似线性可分，使用线性支持向量机，或软间隔支持向量机。

- 引入松弛变量 ξ_i ，使其“可分”，最优化原始问题

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i(w * x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, N \end{aligned}$$

- 求解原始最优化问题的解 w^*, b^* ，得到线性支持向量机，分离超平面为

$$w^* * x + b^* = 0$$

- 分类决策函数为

$$f(x) = \text{sign}(w^* * x + b^*)$$

线性支持向量机的解 w 唯一但 b^* 不一定唯一。

- 对偶问题为

$$\min_a \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j (x_i * x_j) - \sum_{i=1}^N a_i$$

$$\sum_{i=1}^N a_i y_i = 0$$

$$0 \leq a_i \leq C, i = 1, 2, \dots, N$$

- 通过求解对偶问题学习线性可分支持向量机，首先求得最优值 a^* ，然后求最优值 w^*, b^* ，得出分离超平面和分类决策函数。
- 对偶问题的解 a 中满足 $a_i > 0$ 的实例点 x_i 称为支持向量。支持向可在间隔边界上，也可在间隔边界与分离超平面之间，或者在分离超平面误分的一侧。最优分离超平面由支持向量完全决定。
- 线性支持向量机学习等价于最小化二阶范数正则化的合页函数

$$\sum_{i=1}^N [1 - y_i(w * x_i + b)]_+ + \lambda ||w||^2$$

3. 非线性支持向量机

- 对于输入空间中的非线性分类问题，可以通过非线性变换将它转化为某个高维特征空间中的线性分类问题。用核函数替代内积。核函数表示通过一个非线性变换后的两个实例间的内积。

◦ 核函数

- $K(x, z)$ 是一个核函数，或正定核，存在一个从输入空间 X 到特征空间 H 的映射 $\varphi(x)$ ，对任意 x, z 有

$$K(x, z) = \varphi(x) * \varphi(z)$$

对称函数 $K(x, z)$ 为正定核的充要条件为：对任意 x ，对称函数 $K(x, z)$ 对应的 Gram 矩阵是半正定的。

- 所以用核函数表示内积，求得的非线性支持向量机

$$f(x) = \text{sign}(\sum_{i=1}^N a_i^* y_i K(x, x_i) + b^*)$$

4. SMO算法

- SMO算法特点是不断将原二次规划问题分解为只有两个变量的二次规划子问题，并对子问题进行解析求解，直到所有变量满足 KKT 条件为止。

• 仿射函数

- 若 $f(x)$ 满足 $f(x) = ax + b$ ，则称 $f(x)$ 为仿射函数。

• 凸二次规划问题

◦ 凸函数

- 函数任意两点连线处的值大于对应自变量处的函数值。

◦ 凸优化

- 要求目标函数是凸函数，变量所属集合是凸集合优化问题

◦ 函数表示

$$\min_x q(x) = \frac{1}{2} x^T G x + x^T c$$

$$s. t. a_i^T x \geq b_i, i \in \tau$$

- G是Hessian矩阵，T是有限指标集。如果Hessian矩阵是半正定矩阵，则我们说该式是一个凸二次规划。常用拉格朗日求解凸二次规划问题。

- 拉格朗日函数

$$\min f(x)$$

$$s.t. g_I(x) \leq 0$$

$$h_i(x) \leq 0$$

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^n \alpha_i g_i(x) + \sum_{i=1}^n \beta_i h_i(x)$$

- KKT条件:

- 主问题可行

$$g(x) \leq 0$$

- 对偶问题可行

$$\alpha \alpha_i \geq 0$$

- 互补松弛

$$\alpha_i g_i(x) = 0$$