第十一章 对偶理论

第 34 讲 分类模型

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

1. 34.1 分类

1. 34.1 分类

34.1.1 线性判别

给定 R^n 中的两个点集 $\{x_1, \cdots, x_N\}$ 和 $\{y_1, \cdots, y_M\}$ 被某一超平面所分离

$$\begin{aligned} a^T x_i - b &> 0, \quad i = 1, \cdots, N \\ a^T y_i - b &< 0, \quad i = 1, \cdots, M \end{aligned} \tag{1}$$

等价于求解一组 a, b 的线性方程组

$$\begin{aligned} a^T x_i - b &\geq 1, \quad i = 1, \cdots, N \\ a^T y_i - b &\leq -1, \quad i = 1, \cdots, M \end{aligned} \tag{2}$$

34.1.1 线性判别

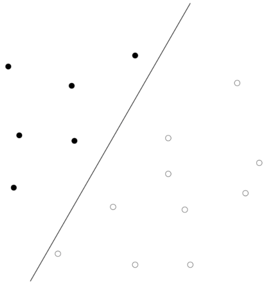

图 1: 点 $x_1, \cdots, x_N$ 由空心圆圈所示，$y_1, \cdots, y_M$ 由实心圆圈所示。两个集合由仿射函数 f 所分类，其 0-水平集（一条直线）分离了它们

34.1.2 鲁棒线性判别

两个超平面之间的欧氏距离

$$
\begin{aligned}
H_1 &= \{\ z\ | a^{\mathrm{T}} z - b = +1\} \\
H_2 &= \{\ z\ | a^{\mathrm{T}} z - b = -1\}
\end{aligned}
$$

为 $t = dist(H_1, H_2)$

通过最大化间隔来分离两个点集

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & a^{\mathrm{T}} x_i - b \geq t, \quad i = 1, \cdots, N \\
& a^{\mathrm{T}} y_i - b \leq -t, \quad i = 1, \cdots, M \\
& \|a\|_2 \leq 1
\end{array} \tag{3}
$$

最优值 t^* 为正，当且仅当两个点集可以线性分离。不等式 $\|a\|_2 \leq 1$ 在最优解处总是紧的，$\|a\|_2 = 1$。图2给出一个简单的几何解释

34.1.2 鲁棒线性判别

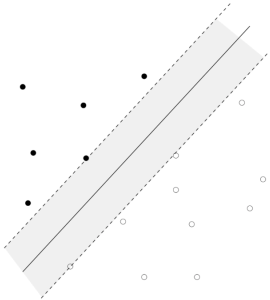

图 2: 通过求解鲁棒线性判别问题 (3)，找到给出两个集合间函数值最大间隙的仿射函数（函数的线性部分有归一化的边界）。几何上，寻找分离两个点集的最宽的点。

34.1.2 鲁棒线性判别

如上图所示，最优值 t^*（即带宽的一半）是两个点集的凸包间距离的一半。（极小化-t 问题）的 Lagrange 为

$$-t+\sum_{i=1}^{N}\mu_i(t+b-a^{\mathrm{T}}x_i)+\sum_{i=1}^{M}\nu_i(t-b+a^{\mathrm{T}}y_i)+\lambda(\|a\|_2-1)$$

在 b 和 t 上极小化得到条件 $a^{\mathrm{T}}\mu=1/2$，$1^{\mathrm{T}}\nu=1/2$。当它们成立时，有

$$\begin{aligned} g(\mu,\nu,\lambda) &= \inf_a\Big(a^{\mathrm{T}}\Big(\sum_{i=1}^{M}\nu_i y_i-\sum_{i=1}^{N}\mu_i x_i\Big)+\lambda\|a\|_2-\lambda\Big) \\ &= \begin{cases} -\lambda & \|\sum_{i=1}^{M}\nu_i y_i-\sum_{i=1}^{N}\mu_i x_i\|_2\le\lambda \\ -\infty & \text{其他情况} \end{cases} \end{aligned}$$

34.1.2 鲁棒线性判别

对偶问题可以写成

$$
\begin{array}{ll}
\text{maximize} & -\|\sum_{i=1}^{M}\nu_i y_i - \sum_{i=1}^{N}\mu_i x_i\|_2 \\
\text{subject to} & \mu \geq 0, \quad 1^{\mathrm{T}}\mu = 1/2 \\
& \nu \geq 1, \quad 1^{\mathrm{T}}\nu = 1/2
\end{array} \tag{4}
$$

可以将 $2\sum_{i=1}^{N}\mu_i x_i$ 解释为 $\{x_1, \cdots, x_N\}$ 的凸包上的一点，而 $2\sum_{i=1}^{M}\nu_i y_i$ 是 $\{y_1, \cdots, y_N\}$ 的凸包上的一点。对偶目标是极小化这两个点之间的（半距离），即寻找两个集合的凸包之间的（半）距离

34.1.3 间隔与支持向量

给定训练样本集 $D = (x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)$， $y_1 \in \{-1, +1\}$，分类学习最基本的想法就是基于训练集 D 在样本空间中找到一个划分超平面、将不同类别的样本分开，但能将训练样本分开的划分超平面可能有很多，我们应该去找到哪一个呢？

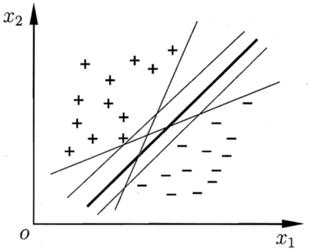

图 3: 存在多个划分超平面将两类训练样本分开

34.1.3 间隔与支持向量

超平面方程：

$$w^{\mathrm{T}}x + b = 0 \tag{5}$$

$w = (w_1, w_2, \cdots, w_d)$ 为法向量，决定了超平面的方向；b 为位移项，决定了超平面与原点之间的距离

样本空间中任一点 x 到超平面 (w, b) 的距离可写为

$$r = \frac{|w^{\mathrm{T}}x + b|}{\|w\|} \tag{6}$$

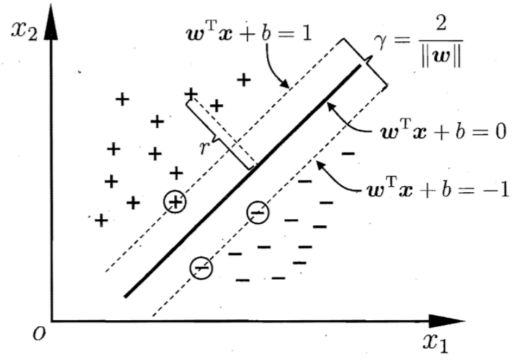

图 4: 支持向量与间隔

34.1.3 间隔与支持向量

欲找到具有“最大间隔”(maximun margin) 的划分超平面，也就是要找到能满足式中约束的参数 w 和 b，使得 γ 最大

$$\max_{w,b} \quad \frac{2}{\|w\|}$$
$$s.t. \quad y_i(w^{\mathrm{T}}x_i + b) \geq 1, \quad i = 1, 2, \cdots, m$$

为了最大化间隔，仅需最大化 $\|w\|^{-1}$，这等价于最小化 $\|w\|^2$。

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$s.t. \quad y_i(w^{\mathrm{T}}x_i + b) \geq 1, \quad i = 1, 2, \cdots, m$$

这是支持向量机（Support vector Machine，简称 SVM）的基本型，本身是一个凸二次规划问题，能直接用现成的优化计算包求解，但有更高效的办法

对偶问题

- 添加拉格朗日乘子 $\alpha \geq 0$，则该问题的拉格朗日函数可写为

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i(1-y_i(w^{\mathrm{T}}x_i+b))$$

- 令 $L(w,b,\alpha)$ 对 w 和 b 的偏导为零可得

$$w=\sum_{i=1}^{m}\alpha_i y_i x_i,\qquad 0=\sum_{i=1}^{m}\alpha_i y_i$$

- 回代可得

$$\max_{\alpha}\quad \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^{\mathrm{T}}x_j$$
$$s.t.\quad \sum_{i=1}^{m}\alpha_i y_i=0,\quad \alpha_i\geq 0,\quad i=1,2,\cdots,m$$

对偶问题

最终模型：

$$f(x) = w^{\mathrm{T}} x + b = \sum_{i=1}^{m} \alpha_i y_i x_i^{\mathrm{T}} x + b$$

KKT 条件

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(x_i) - 1 \geq 0 \\ \alpha_i (y_i f(x_i) - 1) = 0 \end{cases}$$

总有 $\alpha_i = 0$ 或 $y_i(x_i) = 1$

解的稀疏性：训练完成后，大部分的训练样本都不需保留，最终模型仅与支持向量有关，支持向量机因此而得名