1　支持向量机 (SVM)

函数间隔: 对于给定的训练数据集 T 和超平面 (w,b), 定义超平面 (w,b) 关于样本点 (x_i,y_i) 的函数间隔为

$$\hat{\gamma}_i = y_i(wx_i + b)$$

定义超平面 $(w,b$ 关于训练数据集 T 的函数间隔为超平面 (w,b) 关于 T 中所有样本点 (x_i,y_i) 的函数间隔之最小值, 即

$$\hat{\gamma} = \min_{i=1,...,N} \hat{\gamma}_i$$

几何间隔: 对于给定的训练数据集 T 和超平面 (w,b), 定义超平面 (w,b) 关于样本点 (x_i,y_i) 的几何间隔为

$$\gamma_i = y_i(\frac{w}{||w||}x_i + \frac{b}{||w||})$$

定义超平面 $(w,b$ 关于训练数据集 T 的函数间隔为超平面 (w,b) 关于 T 中所有样本点 (x_i,y_i) 的几何间隔之最小值, 即

$$\gamma = \min_{i=1,...,N} \gamma_i$$

1.1　间隔最大化

下面考虑如何求得一个几何间隔最大的分离超平面, 即最大间隔分离超平面. 具体地, 这个问题可以表示为下面的约束最优化问题:

$$\max_{w,b} \gamma$$

$$s.t. \quad y_i(\frac{w}{||w||}x_i + \frac{b}{||w||}) \geq \gamma, i = 1,2,...,N$$

考虑几何间隔和函数间隔的关系 $\gamma = \frac{\hat{\gamma}}{||w||}$, 可将这个问题改写为

$$\max_{w,b} \frac{\hat{\gamma}}{||w||}$$

$$s.t. \quad y_i(wx_i + b) \geq \hat{\gamma}, i = 1,2,...,N$$

函数间隔 $\hat{\gamma}$ 的取值并不影响最优化问题的解. 这样就可以取 $\hat{\gamma} = 1$, 将 $\hat{\gamma} = 1$ 代入上面的最优化问题, 之一到最大化 $\frac{q}{||w||}$ 和最小化 $\frac{1}{2}||w||^2$ 是等价的, 于是就得到下面的线性可分支持向量机学习的最优化问题

$$\min_{w,b} \frac{1}{2}||w||^2$$

$$s.t. \quad y_i(wx_i + b) - 1 \geq 0, i = 1, 2, ..., N$$

这是一个凸二次规划 (convex quadratic programming) 问题.

1.2 学习的对偶算法

首先构建拉格朗日函数 (Lagrange function).

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i y_i(wx_i + b) + \sum_{i=1}^{N} \alpha_i$$

根据拉格朗日对偶性, 原始问题的对偶问题是极大极小问题:

$$\max_{\alpha} \min_{w,b} L(w, b, \alpha)$$

分离超平面可以写成

$$\sum_{i=1}^{N} \alpha_i^* y_i(xx_i) + b^* = 0$$

分类决策函数可以写成

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i^* y_i(xx_i) + b^*)$$

1.3 线性支持向量机

线性不可分的线性支持向量机的学习问题变成如下凸二次规划 (convex quadratic programming) 问题 (原始问题):

$$\min_{w,b,\xi} = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i$$

$$s.t. \quad y_i(wx_i + b) \geq 1 - xi_i, i = 1, 2, ..., N$$

$$\xi_i \geq 0, i = 1, 2, ..., N$$

1.4 核技巧

$$K(x, z) = \phi(x)\phi(z)$$

高斯核函数

$$K(x, z) = \exp(-\frac{||x - z||^2}{2\sigma^2})$$