△白板推导7 核方法

- Kernel Method (思想): 将低维非线性问题转为高维线性问题求解.
- Kernel Trick (计算): 用核函数求内积
- Kernel Function = 非线性转换 + 内积

  ↑来自对偶表示

分类问题:

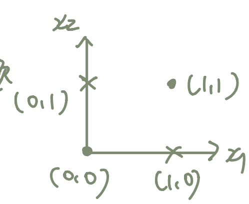
线性可分	线性不可分	
	有一点点错误	严格非线性
PLA (感知机)	Pocket Algorithm	$\phi(x)$+PLA / 多层感知机 (DL)
硬间隔SVM	软间隔SVM	$\phi(x)$+硬间隔SVM = kernel SVM

framework:

$$\underset{\text{input space}}{\mathcal{X}} \xrightarrow[\text{非线性转换}]{\phi(x)} \underset{\text{feature space}}{\mathcal{Z}}$$

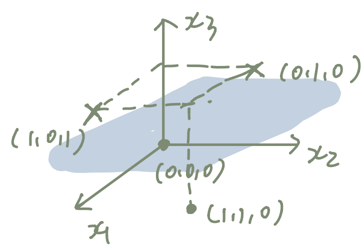
非线性可分 input space → 非线性转换 → feature space 线性可分.

e.g. 异或问题



$x=(x_1,x_2)$ 线性不可分.

$\longrightarrow$ $z=(x_1, x_2, (x_1-x_2)^2)$.



线性可分了.

Cover theorem: 高维空间比低维空间更易线性可分.

硬间隔SVM:

原问题: $\begin{cases} \min\limits_{w,b} \frac{1}{2}w^Tw \\ s.t. \quad y_i(w^Tx_i+b) \geq 1 \quad \text{for all } i=1,\dots,N \end{cases}$

对偶问题: $\begin{cases} \min \frac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boxed{x_i^T x_j} - \sum\limits_{i=1}^{N}\lambda_i \\ s.t. \quad \lambda_i \geq 0 \quad \text{for all } i=1,\dots,N \\ \quad\quad \sum\limits_{i=1}^{N}\lambda_i y_i = 0 \end{cases}$

→ 若经非线性变换, 则为 $\phi(x_i)^T\phi(x_j)$

$\phi(x)$维度可能很高，难求，求内积更难。我们不关心$\phi(x)$具体的值是多少，只关心内积 $\Rightarrow$ 核函数。

通常说核函数指的是正定核函数。

(正定)核函数 定义1：$K: \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$ $\forall x, z \in \mathcal{X}$，有$K(x,z)$

若$\exists \phi: \mathcal{X} \mapsto z$ 且 $\phi \in \mathcal{H}$ s.t. $K(x,z) = \langle\phi(x), \phi(z)\rangle$

则$K(x,z)$为正定核函数。

定义2：$K: \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$ $\forall x, z \in \mathcal{X}$，有$K(x,z)$

若$K(x,z)$满足如下性质：

① 对称性：$K(x,z) = K(z,x)$

② 正定性：$\forall N$个元素 $x_1, \ldots, x_N \in \mathcal{X}$，对应的Gram矩阵半正定

$K = [K(x_i, x_j)]$

则$K(x,z)$为正定核函数。

Hilbert Space

希尔伯特空间$\mathcal{H}$：完备的，可能是无限维的，被赋予内积运算的线性空间

完备的 → 对极限操作封闭

线性空间 → =向量空间，对加法、数乘封闭。

内积运算 →
- 对称性：$\langle f, g\rangle = \langle g, f\rangle$
- 正定性(非负性)：$\langle f, f\rangle \geq 0$，当且仅当$f=0$取等
- 线性：$\langle r_1 f_1 + r_2 f_2, g\rangle = r_1\langle f_1, g\rangle + r_2\langle f_2, g\rangle$

证明定义1和定义2等价：$K(x,z) = \langle\phi(x), \phi(z)\rangle \Leftrightarrow \begin{cases} K(x,z) = K(z,x) \\ \text{Gram矩阵半正定} \end{cases}$

证明"$\Rightarrow$"：

$\because K(x,z) = \langle\phi(x), \phi(z)\rangle$，$K(z,x) = \langle\phi(z), \phi(x)\rangle$

又$\because$内积有对称性 $\Rightarrow$ $\langle\phi(x), \phi(z)\rangle = \langle\phi(z), \phi(x)\rangle$

$\therefore K(x,z) = K(z,x)$。 $K(x,z)$满足对称性。

欲证Gram矩阵 $K = [K(x_i, x_j)]_{N\times N}$ 半正定，

即证 $\forall \alpha \in \mathbb{R}^N$，$\alpha^T K \alpha \geq 0$。

$$\alpha^T K \alpha = (\alpha_1 \ \alpha_2 \cdots \alpha_N)\begin{pmatrix} K_{11} & \cdots & K_{1N} \\ \vdots & \ddots & \vdots \\ K_{N1} & \cdots & K_{NN}\end{pmatrix}\begin{pmatrix}\alpha_1 \\ \vdots \\ \alpha_N\end{pmatrix}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K_{ij} \qquad K_{ij} = K(x_i, x_j)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\langle\phi(x_i), \phi(x_j)\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\phi(x_i)^T\phi(x_j)$$

$$= \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \sum_{j=1}^{N} \alpha_j \phi(x_j)$$

$$= \left\langle \sum_{i=1}^{N} \alpha_i \phi(x_i), \sum_{j=1}^{N} \alpha_j \phi(x_j) \right\rangle$$ ← 这两项的值相同.

$$= \left\| \sum_{i=1}^{N} \alpha_i \phi(x_i) \right\|^2 \geqslant 0$$

∴ K半正定.