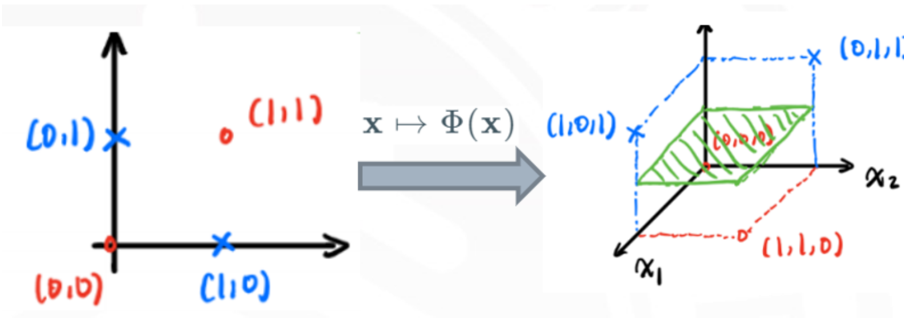
支持向量机2

核技巧



考虑特征转换：$\Phi : \mathbf{R}^p \rightarrow \mathbf{R}^q$

$\mathbf{x} \mapsto \Phi(\mathbf{x})$

则对偶问题可化为：

$$
\begin{aligned}
\min_{\lambda} \quad & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) - \sum_{i=1}^{N} \lambda_i \\
s.t. \quad & \lambda_i \geq 0, i = 1, 2, \ldots, N. \\
& \sum_{i=1}^{N} \lambda_i y_i = 0
\end{aligned}
$$

- 对于该对偶问题，关键是计算$\Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j)$，当$q$很大甚至是无穷大时，这个计算量和存储量都是非常大。
- 核技巧旨在将特征映射和内积这两步运算压缩为一步，即

$$\kappa_\Phi(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j)$$

- 例 二阶多项式转换:

$$\Phi_2(\mathbf{x}) = (1, x_1, x_2, \ldots, x_p, x_1^2, x_1x_2, \ldots, x_1x_p, x_2x_1, x_2^2, \ldots, x_2x_p, \ldots, x_p^2)$$

$$\begin{aligned}
\Phi_2(\mathbf{x})^T\Phi_2(\mathbf{x}') \quad &= 1 + \sum_{i=1}^{p} x_i x_i' + \sum_{i=1}^{p}\sum_{j=1}^{p} x_i x_j x_i' x_j' \\
&= 1 + \sum_{i=1}^{p} x_i x_i' + \sum_{i=1}^{p} x_i x_i' \sum_{j=1}^{p} x_j x_j' \\
&= 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')
\end{aligned}$$

考虑无穷维转换$\Phi: x \mapsto e^{-x^2} \cdot (1, \sqrt{\frac{2}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \ldots)^T$

$$\begin{aligned}\kappa(x_i, x_j) &= \sum_{k=0}^{\infty} (e^{-x_i^2} \sqrt{\frac{2^k}{k!}} x_i^k)(e^{-x_j^2} \sqrt{\frac{2^k}{k!}} x_j^k) \\ &= e^{-x_i^2} e^{-x_j^2} \sum_{k=0}^{\infty} \frac{(2x_i x_j)^k}{k!} \\ &= e^{-x_i^2} e^{-x_j^2} e^{2x_i x_j} \\ &= e^{-(x_i - x_j)^2}\end{aligned}$$

- 高斯核函数(径向基函数)：$\kappa(x, x') = e^{-\gamma\|x - x'\|^2}$
- 核函数：$\kappa: X \times X \to R$
- 正定核函数：对于核函数$\kappa: X \times X \to R$，设$\mathcal{H}$为希尔伯特空间，对于$\forall x, z \in X$, 如果存在一个$\phi: X \to R^q$且$\phi(x) \in \mathcal{H}$, 使得$\kappa(x, z) =< \phi(x), \phi(z) >$，那么称$\kappa(x, z)$为正定核函数.
- 充要条件：$\kappa: X \times X \to R$是正定核函数，当且仅当$\forall x_i \in X (i = 1, 2, \ldots, m)$, 核函数对应的Gram矩阵$K = [\kappa(x_i, x_j)]_{m \times m}$为半正定矩阵。

必要性. 当$\kappa: X \times X \to R$是正定核函数，则存在$\phi: X \to R^q$, 使得$\kappa(x,z) =< \phi(x), \phi(z) >$. 于是对于任意的$x_1, x_2, \ldots, x_m$, 构造对应于Gram矩阵$K = [\kappa(x_i, x_j)]_{m\times m}$,

（1）$K^T = [\kappa(x_i,x_j)]^T = [\kappa(x_j,x_i)] = [\kappa(x_i,x_j)] = K$. 即$K$是对称的。

（2）$\forall \xi = [\xi_1, \xi_2, \ldots, \xi_m]^T \neq 0$,

$$
\begin{aligned}
\xi^T K \xi &= \sum_{i,j=1}^{m} \xi_i \kappa(x_i, x_j) \xi_j \\
&= \sum_{i,j=1}^{m} \xi_i < \phi(x_i), \phi(x_j) > \xi_j \\
&=< \sum_{i=1}^{m} \xi_i \phi(x_i), \sum_{j=1}^{m} \xi_j \phi(x_j) > \\
&= (\sum_{i=1}^{m} \xi_i \phi(x_i))^T (\sum_{j=1}^{m} \xi_j \phi(x_j)) \\
&= \| \sum_{i=1}^{m} \xi_i \phi(x_i)) \|^2 \\
&\geq 0
\end{aligned}
$$

- 定理 若$\kappa(x,z)$是正定核函数，则下列函数也是正定核函数。

$$c_1\kappa_1(x,z)+c_2\kappa_2(x,z), c_1, c_2>0$$

$$\kappa_1(x,z)\kappa_2(x,z)$$

$$f(x)\kappa_1(x,z)f(z)$$

- 常用核函数
  - 线性核$x_i^T x_j$: 有高效实现，不易过拟合，但无法解决非线性可分问题
  - 多项式核$(\beta x_i^T x_j+\theta)^n$: 比线性核更一般，$n$直接描述了被映射空间的复杂度，但参数多，当$n$很大时会导致计算不稳定
  - RBF核$\exp(-\frac{\|x_i-x_j\|^2}{2\sigma^2})$: 只有一个参数，没有计算不稳定问题，但计算慢，过拟合风险大

- 简化版表示定理 优化问题

$$\min_{w} \frac{1}{N} \sum_{i=1}^{N} L(w^T \Phi(\mathbf{x}_i), y_i) + \frac{\alpha}{2} \|w\|^2$$

的解w是样本的线性组合，即 $w = \sum_{i=1}^{N} \lambda_i \Phi(\mathbf{x}_i)$

原版表示定理适用于任意单调递增正则项$\Omega(w)$. 表示定理对损失函数形式没有限制, 这意味着对许多优化问题, 最优解都可以写成样本的线性组合.

更进一步, w⊤ϕ(x) 将可以写成核函数的线性组合

$$w^T \phi(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i \kappa(\mathbf{x}_i, \mathbf{x})$$

通过核函数, 我们可以将线性模型扩展成非线性模型.这启发了一系列基于核函数的学习方法, 统称为核方法。

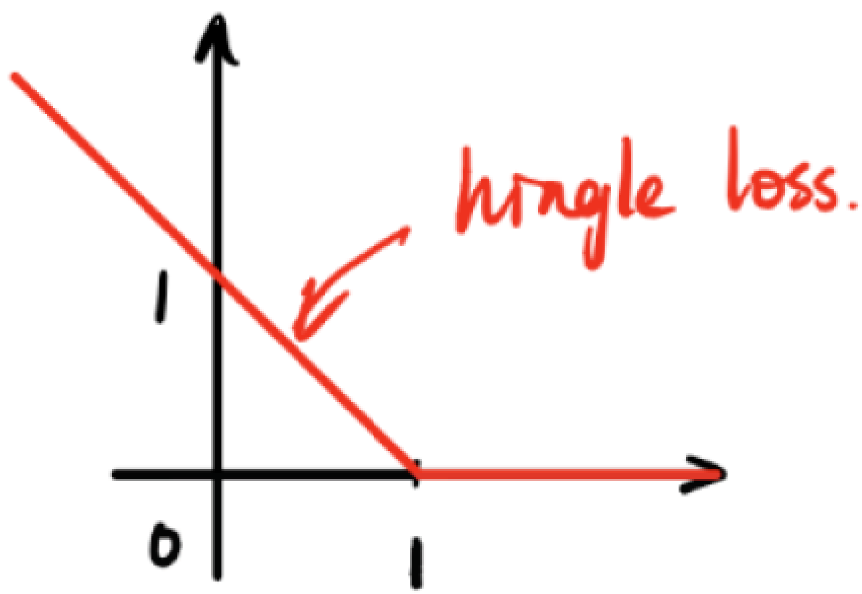
软间隔

- 硬间隔的问题
  - 数据集不一定线性可分
  - 合适的核函数通常很难找到
  - 数据的噪声，强行追求线性可分，容易过拟合
- 软间隔：允许少量样本分类错误

$$\min \frac{1}{2}\mathbf{w}^T\mathbf{w} + loss$$

- $loss$ =错误点的个数= $\sum_{i=1}^{N} I\{y_i(\mathbf{w}^T\mathbf{x}_i + b) < 1\}$

- $loss$ =距离
  - $y_i(w^T\mathbf{x}_i + b) \geq 1, loss = 0$
  - $y_i(w^T\mathbf{x}_i + b) < 1, loss = 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)$
  - 分类的正确率得分：$z = y_i(\mathbf{w}^T\mathbf{x}_i + b)$
  - 这样$loss = \max\{0, 1 - z\}$



$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\max\{0, 1 - y_i(\mathbf{w}^T x_i + b)\} \\
s.t. \quad & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad (i = 1, 2, \ldots, N)
\end{aligned}
$$

- 定义$\xi_i = 1 - y_i(\mathbf{w}^T x_i + b), \xi_i \geq 0$
- 软间隔支持向量机:

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \\
s.t. \quad & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (i = 1, 2, \ldots, N)
\end{aligned}
$$

- C是个可调节参数用于权衡优化间隔和少量样本违背大间隔约束这两个目标.
- 当C比较大时, 我们希望更多的样本满足大间隔约束
- 当C比较小时, 我们允许有一些样本不满足大间隔约束.

- 软间隔SVM的拉格朗日函数为：

$$
\begin{aligned}
L(\mathbf{w}, b, \xi, \alpha, \beta) = \quad & \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \\
& + \sum_{i=1}^{N}\alpha_i(1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)) \\
& + \sum_{i=1}^{N}\beta_i(-\xi_i)
\end{aligned}
$$

- 对偶问题为

$$
\begin{aligned}
\max_{\lambda}\min_{\mathbf{w},b} \quad & L(\mathbf{w}, b, \xi, \alpha, \beta) \\
s.t. \quad & \alpha_i \geq 0, i = 1, 2, \ldots, N, \\
& \beta_i \geq 0, i = 1, 2, \ldots, N.
\end{aligned}
$$

- 软间隔SVM对偶型

$$
\begin{aligned}
\min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \textstyle\sum_{i=1}^{N}\alpha_i \\
s.t. \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C, i = 1, 2, \ldots, N.
\end{aligned}
$$

- 软间隔SVM核对偶型

$$
\begin{aligned}
\min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \textstyle\sum_{i=1}^{N}\alpha_i \\
s.t. \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq \xi_i, i = 1, 2, \ldots, N.
\end{aligned}
$$

形式	优化目标	约束	变量数	约束数
(硬间隔) 基本型	$\frac{1}{2}\boldsymbol{w}^\top\boldsymbol{w}$	$y_i(\boldsymbol{w}^\top\boldsymbol{x}_i+b)\geq 1$	$d+1$	m
(硬间隔) 对偶型	$\frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^\top\boldsymbol{x}_j-\sum_{i=1}^m\alpha_i$	$\sum_{i=1}^m\alpha_i y_i=0,\alpha_i\geq 0$	m	$m+2$
软间隔基本型	$\frac{1}{2}\boldsymbol{w}^\top\boldsymbol{w}+C\sum_{i=1}^m\xi_i$	$y_i(\boldsymbol{w}^\top\boldsymbol{x}_i+b)\geq 1-\xi_i,\xi_i\geq 0$	$m+d+1$	$2m$
软间隔对偶型	$\frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^\top\boldsymbol{x}_j-\sum_{i=1}^m\alpha_i$	$\sum_{i=1}^m\alpha_i y_i=0,0\leq\alpha_i\leq\xi_i$	m	$2m+2$

- 支持向量
- 软间隔SVM的KKT条件
  - 主问题可行：$1-\xi_i-y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b)\leq 0,-\xi_i\leq 0$
  - 对偶问题可行：$\alpha_i\geq 0,\beta_i\geq 0$
  - 互补松弛：$\alpha_i(1-\xi_i-y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b))=0,\beta_i\xi_i=0$
- 软间隔支持向量机中, 支持向量落在最大间隔边界、内部、或被错误分类的样本.
- 支持向量机的参数$(w;b)$仅由支持向量决定,与其他样本无关.

优化方法

- 序列最小化（Sequential Minimal Optimization，SMO）
- 坐标下降：通过循环使用不同坐标方向, 每次固定其他元素, 只沿一个坐标方向进行优化, 以达到目标函数的局部最小

Algorithm 1 坐标下降.

Input: 优化目标 f.

Output: u, 使得 $f(u)$ 最小.

while 不收敛 **do**
    for $i \leftarrow 1$ **to** n **do**
        $u_i \leftarrow \arg\min_{u_i} f(u)$
    end for
end while
return u

- SMO每步的优化目标为:

$$
\begin{aligned}
\min_{\alpha_i,\alpha_j} \quad & \frac{1}{2}(\alpha_i^2 y_i^2 \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) + \alpha_j^2 y_j^2 \phi(\mathbf{x}_j)^T\phi(\mathbf{x}_j) \\
& +2\alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)) - (\alpha_i + \alpha_j) \\
s.t. \quad & \alpha_i y_i + a_j y_j = c \\
& 0 \le \alpha_i \le C \\
& 0 \le \alpha_j \le C
\end{aligned}
$$

其中，$c := -\sum_{k\neq i,j}^{N} \alpha_k y_k$

- SMO 每步的优化目标可等价为对α_i的单变量二次规划问题.