支持向量机

7.3 非线性支持向量机与核函数

7.3.1 核技巧

1. 非线性分类问题

- **非线性分类问题**是指通过利用非线性模型才能很好地进行分类的问题；
- 对于训练集T，如果能用$\mathbb{R}^n$中的一个*超曲面*将正负例正确分开，则称问题为**非线性可分问题**；
- 例子：设原空间为$\mathcal{X} \subset \mathbb{R}^2, x = (x^{(1)}, x^{(2)})^T \in \mathcal{X}$，新空间$\mathcal{Z} \subset \mathbb{R}^2, z = (z^{(1)}, z^{(2)})^T \in \mathcal{Z}$，定义从原空间到新空间的变换（映射）：

$$z = \phi(x) = ((x^{(1)})^2, (x^{(2)})^2)^T \tag{1}$$

原空间中的点相应地变成新空间中的点，原空间中的椭圆变成新空间的直线，且可以将变换后的正负实例点正确分开；

- 用线性分类法求解非线性分类问题：1）使用一个变换将原空间的数据映射到新空间；2）在新空间中用线性分类学习方法从训练数据中学习模型；
- **核技巧**应用到支持向量机：通过一个非线性变换将输入空间（欧式空间$\mathbb{R}^n$或离散集合）对应到一个特征空间（希尔伯特空间$\mathcal{H}$），使得在输入空间中的超曲面模型对应于特征空间的超平面模型；

2. 核函数的定义

- **核函数**定义：设$\mathcal{X}$是输入空间（欧式空间$\mathbb{R}^n$或离散集合），而$\mathcal{H}$为特征空间（希尔伯特空间），如果存在一个从$\mathcal{X}$到$\mathcal{H}$的映射：

$$\phi(x): \mathcal{X} \rightarrow \mathcal{H} \tag{2}$$

使得对所有的$x, z \in \mathcal{X}$，函数$K(x, z)$满足条件：

$$K(x, z) = \phi(x) \cdot \phi(z) \tag{3}$$

则称$K(x, z)$为核函数，$\phi(x)$为映射函数；

- 核技巧的想法：在学习和预测中只定义核函数$K(x, z)$，而不显式地定义映射函数ϕ；通常直接计算$K(x, z)$容易，而通过$\phi(x)$和$\phi(z)$计算$K(x, z)$不容易；
- 对于给定核$K(x, z)$，特征空间$\mathcal{H}$和映射函数ϕ取法不唯一，可以*取不同的特征空间*，在同一特征空间也可以*取不同的映射*；

3. 核技巧在支持向量机中的应用

- 注意到在线性支持向量机的对偶问题中，无论是目标函数还是决策函数，都只涉及输入实例与实例之间的内积；
- 在对偶问题的目标函数中的内积$x_i \cdot x_j$可以用核函数$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$代替，变为

$$W(\alpha) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \tag{4}$$

- 同样分类决策函数中的内积也可以用核函数代替，变成

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x) + b^*\right) \tag{5}$$

- 这样等价于经过映射函数ϕ将原来的输入空间变换到一个新的特征空间中，将输入空间的内积变成特征空间的内积$\phi(x_i) \cdot \phi(x_j)$，在特征空间中学习线性支持向量机；
- 当映射函数是非线性函数时，学习到的含有核函数的支持向量化就是非线性分类模型；
- *学习是隐式*地在特征空间进行的，不需要显式地定义特征空间和映射函数--核技巧；

7.3.2 正定核

- 通常所说的核函数都是**正定核函数**（positive definite kernel function）；
- 假设$K(x,z)$是定义在$\mathcal{X} \times \mathcal{X}$上的对称函数，且对于任意的$x_1, x_2, \ldots, x_m \in \mathcal{X}$，$K(x,z)$关于$x_1, x_2, \ldots, x_m$的Gram矩阵（第一章感知机用于存储训练集内积的矩阵）是半正定的，可以依据函数$K(x,z)$构成一个希尔伯特空间（Hilbert space）：

1. 定义映射ϕ并构成向量空间$\mathcal{S}$
   - 定义映射：$\phi : x \rightarrow K(\cdot, x)$；
   - 对于任意的$x_i \in \mathcal{X}$，$\alpha_i \in \mathbb{R}, i = 1, 2, \ldots, m$，定义线性组合：$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$；
   - 考虑由线性组合为元素的集合$\mathcal{S}$，由于集合*对加法和数乘运算封闭*，所以$\mathcal{S}$构成一个*向量空间*；
2. 在$\mathcal{S}$上定义内积构成内积空间
   - 在$\mathcal{S}$上对于任意的$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i), g(\cdot) = \sum_{j=1}^{l} \beta_j K(\cdot, x_j)$，定义运算$*$：

$$f * g = \sum_{i=1}^{m} \sum_{j=1}^{l} \alpha_i \beta_j K(x_i, z_j) \tag{6}$$

   - 可以证明运算$*$是空间$\mathcal{S}$的内积，即满足下面条件：

$$\begin{aligned}
&(1) \quad (cf) * g = c(f * g), c \in \mathbb{R} \\
&(2) \quad (f + g) * h = f * h + g * h, h \in \mathcal{S} \\
&(3) \quad f * g = g * f \\
&(4) \quad f * f \geq 0 \\
&(5) \quad f * f = 0 \Leftrightarrow f = 0
\end{aligned}$$

- 赋予内积的空间为内积空间，因此$\mathcal{S}$是一个内积空间，$*$是$\mathcal{S}$内积运算，仍然可以写成：

$$f \cdot g=\sum_{i=1}^{m} \sum_{j=1}^{l} \alpha_{i} \beta_{j} K\left(x_{i}, z_{j}\right) \tag{7}$$

3. 将$\mathcal{S}$完备化构成希尔伯特空间
   - 由定义的内积可以得到范数：$\|f\|=\sqrt{f \cdot f}$，因此$\mathcal{S}$是一个**赋范向量空间**；
   - 根据泛函分析理论，对于不完备的赋范向量空间，一定可以使之完备化，得到完备的赋范向量空间$\mathcal{H}$；
   - *一个内积空间，当作为一个赋范向量空间是完备时*，就是希尔伯特空间；
   - 这样的希尔伯特空间$\mathcal{H}$，称为再生核希尔伯特空间（reproducing kernel Hilbert space，RKHS），这是由于核K具有再生性，满足

$$\begin{array}{ll}(1) & K(\cdot, x) \cdot f=f(x) \\ (2) & K(\cdot, x) \cdot K(\cdot, z)=K(x, z)\end{array} \tag{8}$$

     的称为再生核；
4. 正定核的充要条件
   - **正定核的充要条件**：设$K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$是对称函数，则$K(x, z)$为正定核函数的充要条件是对任意$x_{i} \in \mathcal{X}$，$K(x, z)$对应的Gram矩阵：

$$K=\left[K\left(x_{i}, x_{j}\right)\right]_{m \times m} \tag{9}$$

     是半正定矩阵；
   - **正定核的定价定义**：设$\mathcal{X} \subset \mathbb{R}^{n}$，$K(x, z)$是定义在$\mathcal{X} \times \mathcal{X}$上的对称函数，如果对于任意$x_{i} \in \mathcal{X}$，$K(x, z)$对于的Gram矩阵是半正定矩阵，则称$K(x, z)$是正定核；
   - 该定义在构造核函数时很有用，但对于一个具体函数$K(x, z)$，检验是否为正定核函数并不容易（因为要对任意有限输入集验证K对于的Gram矩阵是否半正定），因此在实际中往往应用已有的核函数；

7.3.3 常用核函数

1. 多项式核函数（polynomial kernel function）

$$K(x, z)=(x \cdot z+1)^{p} \tag{10}$$

- 对应的支持向量机是一个p次多项式分类器，分类决策函数成为：

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{N}\alpha_i^* y_i (x_i\cdot x+1)^p+b^*\right) \tag{11}$$

2. 高斯核函数（Gaussian kernel function）

$$K(x,z)=\exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \tag{12}$$

- 对于的支持向量机是高斯径向基函数（radial basis function）分类器，分类决策函数成为：

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{N}\alpha_i^* y_i \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)+b^*\right) \tag{13}$$

3. 字符串核函数（string kernel function）

- 核函数不仅可以定义在欧式空间上，还可以是离散数据的集合（字符串核定义在字符串集合上，用于文本分类、信息检索、生物信息等）；
- 考虑一个有限字符表Σ，字符串s是从中取出的有限个字符的序列（包括空字符串），字符串s的长度用$|s|$表示，元素记作$s(1),s(2),\ldots,s(|s|)$，两个字符串s和t的连接记作st，所有长度为n的字符串的集合记作Σ^n，所有字符串的集合记作$\Sigma^*=\bigcup_{n=0}^{\infty}\Sigma^n$；
- 考虑字符串s的子串u，给定一个指标序列$i=(i_1,i_2,\ldots,i_{|u|}),1\le i_1<i_2<\cdots<i_{|u|}\le|S|$，$s$的子串定义为$u=s(i)=s(i_1)\cdots s(i_{|u|})$，其长度记作$l(i)=i_{|u|}-i_1+1$，如果$i$是连续的，则$l(i)=|u|$，否则$l(i)>|u|$；
- 假设$\mathcal{S}$是长度大于或等于n 字符串集合，s是$\mathcal{S}$的元素，现在建立字符串集合$\mathcal{S}$到特征空间$\mathcal{H}_n=\mathbb{R}^{\Sigma^n}$的映射$\phi_n(s)$，$\mathbb{R}^{\Sigma^n}$表示定义在$\Sigma^n$上的实数空间，其每一维对应一个字符串$u\in\Sigma^n$，映射将字符串$s$对应于空间$\mathbb{R}^{\Sigma^n}$的一个向量，其在$u$维上的取值为

$$[\phi_n(s)]_u=\sum_{i:s(i)=u}\lambda^{l(i)} \tag{14}$$

其中$0\le\lambda\le 1$是一个衰减参数，$l(i)$表示字符串i 长度，求和在s中所有与u相同的子串上进行；

- 两个字符串s和t上的字符串核函数是基于映射ϕ_n的特征空间中的内积：

$$k_n(s,t) = \sum_{u\in\Sigma^n} [\phi_n(s)]_u [\phi_n(t)]_u = \sum_{u\in\Sigma^n} \sum_{(i,j):s(i)=t(j)=u} \lambda^{l(i)} \lambda^{l(j)} \tag{15}$$

字符串核函数$k_n(s,t)$给出了两字符串中长度等于n 所有子串组成的特征向量的余弦相似度（cosine similarity）；

- 直观上，两个字符串相同的子串越多，就越相似，字符串核函数的值就越大；

7.3.4 非线性支持向量机

- 将线性支持向量机扩展到非线性支持向量机，只需将线性支持向量机*对偶形式中的内积换成核函数*;
- **非线性支持向量机**定义：从非线性分类训练集，通过核函数与软间隔最大化，或凸二次规划学习得到的分类决策函数

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x) + b^*\right) \tag{16}$$

称为非线性支持向量机，$K(x,z)$是正定核函数;

- **非线性支持向量机学习算法**：
  1. **输入**训练数据集T;
  2. 选取适当的核函数$K(x,z)$和适当的参数C，构造并求解最优化问题，得到最优解$\alpha^* = (\alpha_i^*, \alpha_2^*, \ldots, \alpha_N^*)^T$；

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j K(x_i,x_j) - \sum_{i=1}^{N}\alpha_i \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{N}\alpha_i y_i = 0 \\ & 0 \le \alpha_i \le C, \qquad i=1,2,\ldots,N \end{aligned}$$

  3. 选择α^*中的一个正分量$0 < \alpha_j^* < C$，计算

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x_j) \tag{17}$$

  4. 构造决策函数

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x_i, x) + b^*\right) \tag{18}$$