

第7章 支持向量机

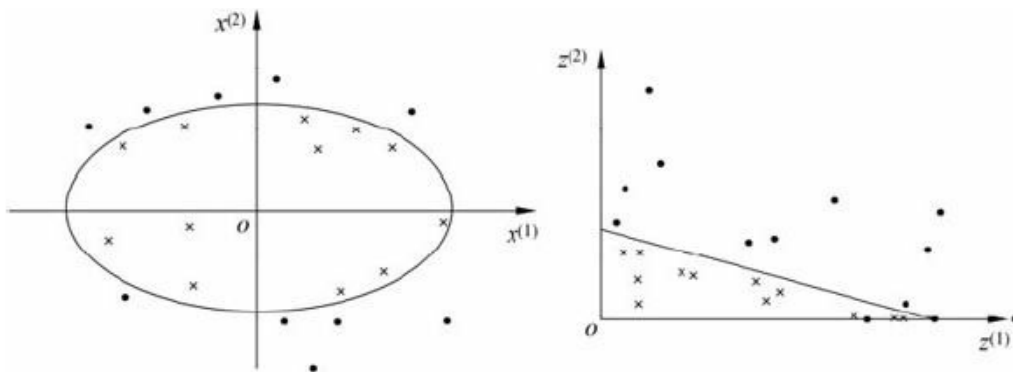
- 上节笔记完成了[线性支持向量机](#)
- 本节讲述非线性支持向量机

7.3 非线性支持向量机与核函数

7.3.1 核技巧

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



- 如上图，无法用直线（线性模型）将正负实例正确分开，但可以用一条椭圆曲线（非线性模型）将它们正确分开
- 非线性可分问题：利用**超曲面**将正负例数据正确分离
- **非线性问题往往不好求解**，所以希望能用解**线性分类**问题的方法解决这个问题
 - 采取的方法是进行一个**非线性变换**，将非线性问题变换为线性问题，通过解**变换后的线性问题**的方法求解**原来的非线性问题**
 - 非线性变化**例子**：

设原空间为 $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{x} = (x^{(1)}, x^{(2)})^T \in \mathbf{x}$ ，新空间为 $\mathbf{z} \in \mathbb{R}^2$, $\mathbf{z} = (z^{(1)}, z^{(2)})^T \in \mathbf{z}$ ，定义从原空间到新空间的变换（映射）：

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

经过变换 $\mathbf{z} = \phi(\mathbf{x})$ ，原空间 $\mathbf{x} \in \mathbb{R}^2$ 变换为新空间 $\mathbf{z} \in \mathbb{R}^2$ ，原空间中的点相应地变换为新空间中的点，**原空间中的椭圆**

$$w_1 (x^{(1)})^2 + w_2 (x^{(2)})^2 + b = 0$$

变换成为**新空间中的直线**

$$w_1 z^{(1)} + w_2 z^{(2)} + b = 0$$

在变换后的新空间里，直线 $w_1 z^{(1)} + w_2 z^{(2)} + b = 0$ 可以将变换后的正负实例点正确分开。这样，原空间的**非线性可分问题**就变成了新空间的**线性可分问题**。

- 可见，利用线性分类方法求解非线性分类有两个步骤：
 - 首先使用一个**变换**将原空间的数据**映射**到新空间
 - 在新空间里用**线性分类**学习方法从训练数据中学习分类模型
- 总结，**核技巧**就类似于上述的变换：
 - 通过一个**非线性变换**将输入空间（欧氏空间 \mathbb{R}^n 或离散集合）对应于一个特征空间（希尔伯特空间 \mathbb{H} ）

- 使得在输入空间 R^n 中的**超曲面**模型对应于特征空间 \mathbb{H} 中的**超平面**模型（支持向量机）

核函数的定义

• 核函数定义 7.6

定义 7.6（核函数） 设 \mathcal{X} 是输入空间（欧氏空间 \mathbb{R}^n 的子集或离散集合），又设 \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

$$\phi(x): \mathcal{X} \rightarrow \mathcal{H} \quad (7.65)$$

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

$$K(x, z) = \phi(x) \cdot \phi(z) \quad (7.66)$$

则称 $K(x, z)$ 为核函数， $\phi(x)$ 为映射函数，式中 $\phi(x) \cdot \phi(z)$ 为 $\phi(x)$ 和 $\phi(z)$ 的内积。

- 核技巧的想法是，在学习与预测中只定义核函数 $K(x, z)$ ，而不显式地定义映射函数 ϕ
 - 直接计算 $K(x, z)$ 比较容易，而通过 $\phi(x)$ 和 $\phi(z)$ 计算 $K(x, z)$ 并不容易
 - ϕ 是输入空间 R^n 到特征空间 \mathbb{H} 的映射，特征空间一般是高维的，甚至是无穷维的，所以先计算映射的话可能在高维度进行，**计算量非常大**
 - 对于给定的核 $K(x, z)$ ，特征空间和映射函数 ϕ 的取法并不唯一，可以取不同的特征空间，即便是在同一特征空间里也可以取不同的映射
 - 举例可见：[椭圆数据空间维度转换](#)

核技巧在支持向量机中的应用

- 有了核技巧，可以对线性支持向量机中的内积 $x_i \cdot x_j$ 用核函数 $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 代替，所以对偶问题的目标函数成为：

$$W(\alpha) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \quad (7.67)$$

- 分类决策函数中的内积也可以用核函数代替，而分类决策函数式成为

$$f(x) = \text{sign} \left(\sum_{i=1}^{N_s} a_i^* y_i \phi(x_i) \cdot \phi(x) + b^* \right) = \text{sign} \left(\sum_{i=1}^{N_s} a_i^* y_i K(x_i, x) + b^* \right) \quad (7.68)$$

- 这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，将输入空间中的内积 $x_i \cdot x_j$ 变换为特征空间中的内积 $\phi(x_i) \cdot \phi(x_j)$ ，在新的特征空间里从训练样本中学习**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机是**非线性分类模型**
- 所以，在核函数 $K(x, z)$ 给定的条件下，可以利用**解线性分类问题的方法求解非线性分类问题的支持向量机**
 - **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，减少大量的计算

7.3.3 常用核函数

- 多项式核函数（polynomial kernel function）

$$K(x, z) = (x \cdot z + 1)^p \quad (7.88)$$

- 对应的支持向量机是一个 p 次多项式分类器。在此情形下，分类决策函数成为

$$f(x) = \text{sign} \left(\sum_{i=1}^{N_s} a_i^* y_i (x_i \cdot x + 1)^p + b^* \right) \quad (7.89)$$

- 高斯核函数（Gaussian kernel function）

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (7.90)$$

- 对应的支持向量机是**高斯径向基函数**（radial basis function, RBF）分类器。在此情形下，分类决策函数成为

$$f(x) = \text{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) + b^*\right) \quad (7.91)$$

- 字符串核函数（string kernel function）

参考资料

- [李航《统计学习方法》第七章](#)