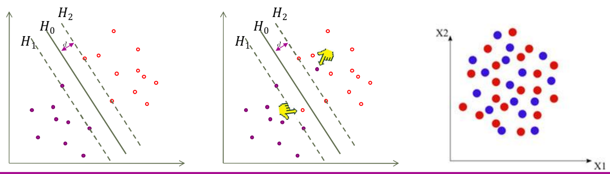
非线性可分问题

非线性可分：不能用一个超平面对数据进行分隔；

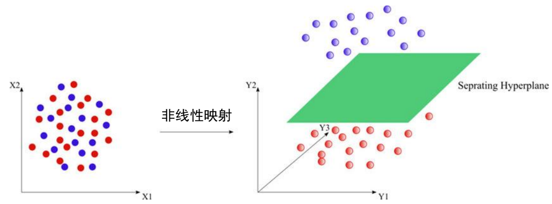
① 离群点导致的非线性可分: 我们可以把非线性问题转化为“线性可分”问题，所以我们还是使用的“线性模型”来解决非线性可分问题

② 数据本质上是非线性可分：无法通过放松分隔条件来进行“变通”



非线性可分问题

- 非线性问题往往不好求解
- 能不能转化为线性分类问题？



非线性可分问题

- 方法: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。

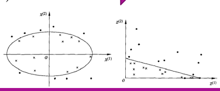
原空间：

$$\mathcal{X} \subset R^2, x=(x^{(1)},x^{(2)})^T \in \mathcal{X}$$

新空间：

$$\mathcal{Z} \subset R^2, z=(z^{(1)},z^{(2)})^T \in \mathcal{Z} \qquad z=\phi(x)=((x^{(1)})^2,(x^{(2)})^2)^T$$

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



线性问题

核函数在支持向量机的应用

- 避免把低维度的数据转化到高维度的空间中，然后再去寻找线性分割平面
- 线性支持向量机对偶问题中，无论是目标函数还是决策函数都只涉及输入实例和实例之间的内积。

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

- 核函数：直接定义高纬空间中的内积

核函数在支持向量机的应用

- 注意到：
- 线性支持向量机对偶问题中，无论是目标函数还是决策函数都只涉及输入实例和实例之间的内积。

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

- 目标函数中的内积 $x_i \cdot x_j$ 用核函数 $K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$

代替，目标函数：

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

非线性支持向量机与核函数

- 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

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

- 直接计算$K(x, z)$ 比较容易，
- 而通过$\phi(x)$ 和$\phi(z)$ 计算并不容易。

非线性支持向量机与核函数

- 核函数定义：
- 设X是输入空间(欧氏空间R^n的子集或离散集合)，又设H为特征空间(希尔伯特空间)，如果存在一个从X到H的映射$\phi(x): \mathcal{X} \to \mathcal{H}$
- 使得对所有$x, z \in \mathcal{X}$
- 函数$K(x,z)$满足条件$K(x,z) = \phi(x) \cdot \phi(z)$
- 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

非线性支持向量机与核函数

- 例：

假设输入空间是R^2，核函数是 $K(x,z) = (x \cdot z)^2$ ，试找出其相关的特征空间H和映射$\phi(x): R^2 \to \mathcal{H}$

解：

取特征空间$\mathcal{H} = R^3$ ，记$x = (x^{(1)}, x^{(2)})^T, z = (z^{(1)}, z^{(2)})^T$

$$(x \cdot z)^2 = (x^{(1)}z^{(1)} + x^{(2)}z^{(2)})^2 = (x^{(1)}z^{(1)})^2 + 2x^{(1)}z^{(1)}x^{(2)}z^{(2)} + (x^{(2)}z^{(2)})^2$$

可以取： $\phi(x) = ((x^{(1)})^2, \sqrt{2}x^{(1)}x^{(2)}, (x^{(2)})^2)^T$

容易验证：$\phi(x) \cdot \phi(z) = (x \cdot z)^2 = K(x,z)$

非线性支持向量机与核函数

- 例：

假设输入空间是R^2，核函数是 $K(x,z) = (x \cdot z)^2$ ，试找出其相关的特征空间H和映射$\phi(x): R^2 \to \mathcal{H}$

解：

同样：

$$\phi(x) = \frac{1}{\sqrt{2}}((x^{(1)})^2 - (x^{(2)})^2, 2x^{(1)}x^{(2)}, (x^{(1)})^2 + (x^{(2)})^2)^T$$

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

都满足条件。

正定核

- 问题：
  - 已知映射函数 ϕ ，可以通过$\phi(x)$ 和$\phi(z)$ 的内积求得核函数$K(x,z)$。
  - 直接给定的函数$K(x,z)$，满足什么条件才能成为正定核函数？
  - 正定核的充要条件
  - 设K：$\mathcal{X} \times \mathcal{X} \to R$ ，是对称函数，则$K(x,z)$为正定核函数的充要条件是对任意 $\in \mathcal{X},\ i = 1,2,...,m$ ，$K(x,z)$对应的Gram矩阵

$$K = [K(x_i, x_j)]_{m \times m}$$

是半正定的。

正定核

- 正定核的等价定义
- 设$\mathcal{X} \subset R^n$ ，是$K(x,z)$是定义在$\mathcal{X} \times \mathcal{X}$ 对称函数，如果对任意的$x_i \in \mathcal{X}, i = 1,2,...,m$ ，$K(x,z)$对应的Gram矩阵

$$K = [K(x_i, x_j)]_{m \times m}$$

是半正定的，则称$K(x,z)$为正定核。

- 这一定义在构造核函数时很有用。但对于一个具体函数$K(x,z)$ 来说，检验它是否为正定核函数并不容易。
- 因为要求对任意有限输入集 $\{x_1, x_2, ..., x_m\}$ 验证K对应的Gram矩阵是否为半正定的。
- 在实际问题中往往应用已有的核函数。

常用核函数

1、多项式核函数(Polynomial kernel function)

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

- 对应的支持向量机为P次多项式分类器，分类决策函数：

$$f(x) = sign(\sum_{i=1}^{N_s} \alpha_i^* y_i (x_i \cdot x + 1)^p + b^*)$$

2、高斯核函数(Gaussian Kernel Function)

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

- 决策函数：

$$f(x) = sign(\sum_{i=1}^{N_s} \alpha_i^* y_i \exp(-\frac{\|x - z\|^2}{2\sigma^2}) + b^*)$$

常用核函数

3、字符串核函数：

- 考虑一个有限字符表∑，字符串s是从∑中取出的有限字符的序列，包括空字符。s长度用|s|表示，元素记作$s(1)s(2)...s(|s|)$ 。两串字符s和t的连接记作st。所有长度为n的字符串的集合记作 ，所有字符串记作

$$\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$$

- 考虑s的子串u，给定一个指标序列

$$i = (i_1, i_2, ...i_{|u|}), 1 \le i_1 < i_2 < ... < i_{|u|} \le |s|$$

- u定义为 $u = s(i) = s(i_1)s(i_2)...s(i_{|u|})$，长度记作$l(i) = i_{|u|} - i_i + 1$

如果i是连续的，则 $l(i) = |u|$， 否则$l(i) > |u|$

核函数的一些性质

核函数的和仍是核函数

$$K(x, y) = K_1(x, y) + K_2(x, y)$$

核函数的积，仍是核函数

$$K(x, y) = K_1(x, y)K_2(x, y)$$

核函数的两端放缩，仍是核函数

$$K(x, y) = f(x)K_1(x, y)f(y)$$

非线性支持向量机学习算法

- 输入：线性不可分训练数据集$T = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$

$$x_i \in \mathcal{X} = R^n, y_i \in \mathcal{Y} = \{-1, +1\}, i = 1, 2, ..., N$$

- 输出：分类决策函数

1、选取适当的核函数和参数C，构造最优化问题：

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

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$0 \le \alpha_i \le C, \ i = 1, 2, ..., N$$

- 求得最优解： $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_l^*)^T$

非线性支持向量机学习算法

2、并选择α*，适合条件$0 < \alpha_j^* < C$ ，计算

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

3、构造决策函数

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

当$K(x, z)$是正定核函数时，优化问题是凸二次规划问题，解是存在的。