

核函数

预备知识

1. 线性可分支持向量机的分类超平面为： $\mathbf{w}^* \mathbf{x} + b^* = 0$ ，其相应的分类决策函数为

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^* \mathbf{x} + b^*)$$

2. 根据对偶问题可知： $\mathbf{w}^* = \sum_{i=1}^N \alpha^* y_i \mathbf{x}_i, b^* = y_j - \sum_{i=1}^N y_i \alpha_i^* (\mathbf{x}_i \mathbf{x}_j)$ ，其中 y_j 对应的样本为支持向量。

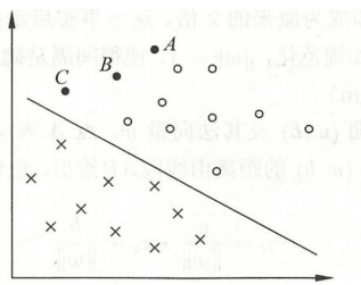
3. 根据1,2可知，分类决策函数为：

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N \alpha^* y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b\right) \quad (4.1)$$

正文

线性可分与线性不可分：

以下图所示的二维特征空间中的分类问题为例，在这种情况下，找到能将两类数据正确划分并且间隔最大的直线即可，所以为线性可分：



但是，现实问题常常存入如下线性不可分的情况，如下图所示：

![[4-2 线性不可分二分类图片](images/4-2 线性不可分二分类图片.png)]

判定边界的模型可能是： $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \dots$ 的形式。用新的特征 f 来替换模型中的每一项，令： $f_1 = x_1, f_2 = x_2, f_3 = x_1 x_2 \dots$ ，也就是将低维特征映射到高维空间，得到： $h_{\theta}(x) = \theta_0 + \theta_1 f_1 + \dots + \theta_n f_n$ 。

上面我们根据多项式经验，构造了新的特征 f 。那么，在给定样本量 m ，原有特征数为 k 的情况下，新特征该怎么构造呢？在[课时97-核函数2]([吴恩达机器学习 - 网易云课堂](#))中，我们构造 $m+1$ 个新特征。对于给定的样本 $(x^{(i)}, y^{(i)})$ （与《统计学习方法》的表达有些差异，第 i 个样本点表示为 x_i, y_i ）其新特征为：

!4-4 新特征(images/4-4 新特征.png)

新特征为给定样本与训练集中所有样本的距离。

核技巧

由上可知，将低维特征映射到高维空间，则可以处理线性不可分的情况。但是将低维映射到高维 $\mathbf{x} \rightarrow \phi(\mathbf{x})$ ，存在一定计算量，如 $f_1 = x_1, f_2 = x_2, f_3 = x_1 x_2 \dots$ 。此外， $\langle \mathbf{x}_i, \mathbf{x} \rangle \rightarrow \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ 计算量也会增大。根据式4.1可知，SVM决策函数最终是要得到 $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ ，如果有函数 $K(\mathbf{x}_i, \mathbf{x}) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ ，这样就不用构造 $\phi(\mathbf{x})$ ，也避免直接在高维空间中进行计算，而结果却是等价的！这时就可以用核函数，直接采用《统计学习方法》上的核函数定义：

定义 7.6 (核函数) 设 \mathcal{X} 是输入空间（欧氏空间 \mathbf{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。

常见核函数

!4-5 常用核函数(images/4-5 常用核函数.png)

其中，高斯核函数通过泰勒展开可以将 \mathbf{x} 映射到无穷维。如果 σ 很大，高次特征上的权重衰减得很快，所以实际上相当于一个低维的子空间；相反，如果 σ 很小，则可以将任意的数据映射为线性可分，但存在过拟合问题。不过，总的来说，通过调控参数，高斯核实际上具有相当高的灵活性，也是使用最广泛的核函数之一。下面总结下高斯核函数：

- σ 较大时，可能会导致低方差，高偏差；
- σ 较小时，可能会导致低偏差，高方差。