

Kernel Methods

FrankZhou-jun*

2019 年 9 月 18 日

目录

1 超平面	1
1.1 平面方程	1
1.2 Hyperplane	3
2 特征值与特征向量	4
3 核函数	6

1 超平面

1.1 平面方程

首先理解一下什么叫平面方程，常用的有点法式、截距式、三点式¹，

如果知道平面上的一个点，以及垂直于该平面的法线，我们就可以得到这平面啦，假设这个平面在三维空间下，依据点构成线，线构成面，面构成体，其实线、面、体都是由点构成构成的，我们求的平面方程，其实就是求满足条件下所有点。由于我们通常是在三维坐标下考虑的平面，所以平面类的点是三维的，当在大于三维空间的情况下，比如四维空间、五维空间、十维空间等，平面内的点对应维数为四维、五维、十维等。之前看到的某个帖子，当我们考虑的空间维数大于三维，所以叫超平面。

继续前面的例子哈，现在我们在三维空间下考虑平面的问题，如果我们知道平面内的点 $P_1(x_1, y_1, z_1)$ 和垂直于该平面的法线 \vec{n} ，我们就可以通过点法式求该平面啦。可以先看看图1

*研究方向：信号处理，机械故障诊断，深度学习，强化学习，邮箱:zhoujun14@yeah.net

¹高等数学

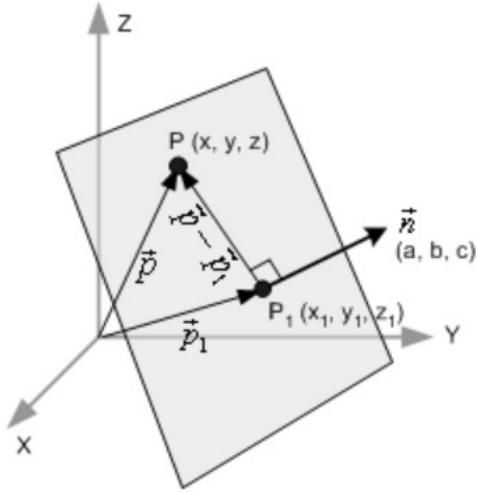


图 1: 点法式

根据“垂直这一条件”，我们可以找到该平面内任意一点 $P(x, y, z)$ 与点 $P_1(x_1, y_1, z_1)$ 构成的向量 $\overrightarrow{P_1P}$ 与法向量 $\vec{n}(a, b, c)$ 相乘为零。

$$\begin{aligned}
 \overrightarrow{P_1P} \cdot \vec{n} &= 0 \\
 (x - x_1, y - y_1, z - z_1) \cdot (a, b, c) &= 0 \\
 a(x - x_1) + b(y - y_1) + c(z - z_1) &= 0 \\
 ax + by + cz - ax_1 - by_1 - cz_1 &= 0
 \end{aligned} \tag{1}$$

法向量 \vec{n} 与点 $P_1(x_1, y_1, z_1)$ 已知，令 $D = -(ax_1 + by_1 + cz_1)$ ，则得到公式如下：

$$ax + by + cz + D = 0 \tag{2}$$

这样就可以得到对应的平面方程啦，可以看到该平面是由满足条件的所有点 $P(x, y, z)$ 组成。可以看大点构成面哈，我们一般叫求平面方程，其实求的是点。

核函数的意义就是，大白话来说就是将低维空间上的点映射到高维空间，是在低维空间不易区分的特征可以在高维上区分，如图2

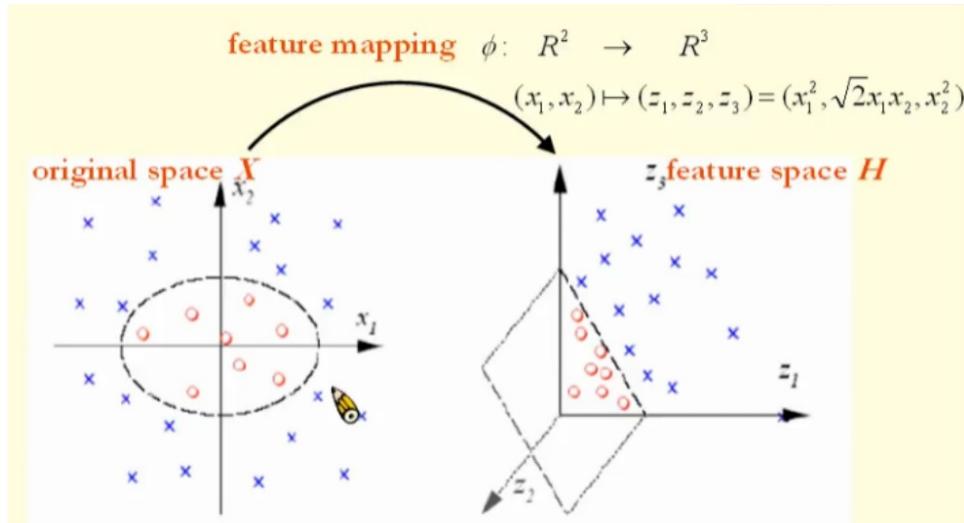


图 2: 超平面

可以看到在低维空间需要一个椭圆才能将两个类分开，实际情况可能比较复杂，简单的椭圆不可能完全分开，可能需要弯弯曲曲的才行。而在右图可以看到用以个超平面就可以将两类区分，所以核函数存在具有明显的价值，使线性

不可分的数据转为线性可分数据²。。

1.2 Hyperplane

讲了这么久终于到超平面了，图2可以看到在三维空间下的超平面，超平面是 n 维欧氏空间中余维度等于一的线性子空间³，也就是必须是 (n-1) 维度。在维度大于三时，划分分类的那个东东就不仅仅是二维的一个平面的，可能是三维，也可能是四维的，所以叫超平面，其维度等于环境空间的维度-1。

为啥超平面的自由度比空间维度小 1，举个例子，三维空间的超平面 $ax + by + cz + D = 0, a, b, z$ 中至少一个不为零，假设 $b \neq 0$, 则

$$y = -\frac{a}{b}x - \frac{c}{b}z - \frac{D}{b} \quad (3)$$

可以看到在三维空间中的超平面，当有两个维度是自由的时候，另一个维度就不自由了，即大小确定，所以二维空间的超平面的维度为 2.

n 维空间下的超平面定义如下：

$$\mathbf{W}^T \mathbf{x} + d = 0 \quad (4)$$

其中 \mathbf{W}, \mathbf{x} 的维度都为 n

$$\mathbf{w} = (w_1, w_2, \dots, w_n)^T \quad (5)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T \quad (6)$$

根据前面的推理可以得到，若 \vec{n} 为超平面的法向量，则超平面方程可以表示为

$$\vec{n} \cdot (\mathbf{P} - \mathbf{P}_1) = 0 \quad (7)$$

超平面是怎么分类的呢，很简单的一个知识点，就是就点到平面的距离，哎，高中时候就学过了，这会到忘了，没事，我们在复习一遍。

空间中 \mathbf{P}_2 到超平面的距离 D 等于超平面上的点 \mathbf{P}_1 指向 \mathbf{P}_2 构成的向量 $\overrightarrow{\mathbf{P}_1 \mathbf{P}_2}$ 在该平面法向量 \vec{n} 上投影的距离长度。参考图3

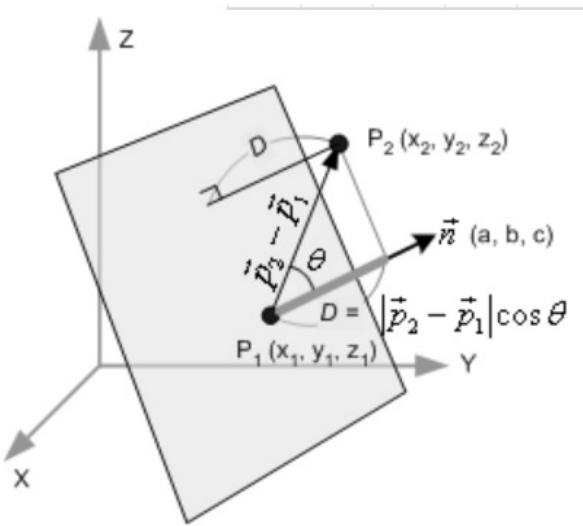


图 3: 点到平面的距离

距离 D 的大小表示为

$$D = |\overrightarrow{\mathbf{P}_1 \mathbf{P}_2}| \cos \theta \quad (8)$$

²<https://www.quora.com/What-are-kernels-in-machine-learning-and-SVM-and-why-do-we-need-them/answer/Lili-Jiang?srqid=o0gT>

³百度

而我们右知道两向量的夹角公式，利用点 P_1 和点 P_2 构成的向量 $\overrightarrow{P_1P_2}$ 与法向量 \vec{n} 计算角度 $\cos \theta$

$$\cos \theta = \frac{\vec{n} \cdot \overrightarrow{P_1P_2}}{|\vec{n}| |\overrightarrow{P_1P_2}|} \quad (9)$$

$$\begin{aligned} D &= |\overrightarrow{P_1P_2}| \cos \theta \\ &= |\overrightarrow{P_1P_2}| * \frac{\vec{n} \cdot \overrightarrow{P_1P_2}}{|\vec{n}| |\overrightarrow{P_1P_2}|} \\ &= \frac{\vec{n} \cdot \overrightarrow{P_1P_2}}{|\vec{n}|} \end{aligned} \quad (10)$$

可以看到化简后方程的分子部分就为环境空间的超平面方程，则该环境空间中的任意一点的距离到超平面的距离表示为

$$D = \frac{\mathbf{W}^T \mathbf{x} + d}{|\vec{n}|} \quad (11)$$

$$T_\sigma = \begin{cases} \text{above plane, } \mathbf{W}^T \mathbf{x} + d > 0 \\ \text{in plane, } \mathbf{W}^T \mathbf{x} + d = 0 \\ \text{under plane, } \mathbf{W}^T \mathbf{x} + d < 0 \end{cases} \quad (12)$$

2 特征值与特征向量

先来讲讲定义：设 A 为 n 阶方阵，若存在 n 阶非零向量 \vec{x} 使得

$$A\vec{x} = \lambda\vec{x} \quad (13)$$

则称 λ 为矩阵 A 的特征值， \vec{x} 为对应特征值下矩阵 A 的特征向量。一个矩阵乘以向量可以表示为在这向量方向上进行伸缩 λ ，有什么用呢，以后再不出。

先上两张图

假设矩阵 A 不变， \vec{x} 变化，如下图⁴

再看看改变矩阵某一列，如何变化。

⁴<http://mini.eastday.com/bdmip/180328092726628.html>

可以看到向量可以经过矩阵“作用”后，沿着该向量方向伸缩。离开大学这么久，感觉连求矩阵的特征值和特征向量都不会了，来个例题先算算：

求 $A = \begin{vmatrix} 1 & -2 \\ 1 & 4 \end{vmatrix}$ 的特征值和特征向量。

$$|\lambda E - A| \vec{x} = 0 \quad (14)$$

方程 $|\lambda E - A| = 0$ 为矩阵 A 的特征方程。

定理：n 阶矩阵 A 的 n 个特征值就是其特征方程的 n 个根；而 A 的属于特征值的特征向量就是其次线性方程的 $|\lambda E - A| \vec{x} = 0$ 非零解。

先求 λ

$$\begin{aligned} |\lambda E - A| &= 0 \\ \begin{vmatrix} \lambda - 1 & 2 \\ -1 & \lambda - 4 \end{vmatrix} &= 0 \\ (\lambda - 1)(\lambda - 4) + 2 &= 0 \end{aligned} \quad (15)$$

求得解 $\lambda_1 = 2, \lambda_2 = 3$ ，再将 λ 分别代入对应的线性方程求解，即得到对应 λ 下的特征向量。

(1) 求出 λ_1 对应的特征向量

$$\begin{vmatrix} 1 & 2 \\ -1 & -2 \end{vmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \quad (16)$$

$$x_1 = 2x_2$$

即，求得基础解析

$$p_1 = k_1 \begin{bmatrix} -2 \\ 1 \end{bmatrix}, k_1 \neq 0 \quad (17)$$

因为满足 $x_1 = 2x_2$ 的条件很多，得到的特征向量不唯一，但方向一样。

(2) 求出 λ_2 对应的特征向量

$$\begin{vmatrix} 2 & 2 \\ -1 & -1 \end{vmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \quad (18)$$

$$x_1 = -x_2$$

即，求得基础解析

$$p_1 = k_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}, k_2 \neq 0 \quad (19)$$

3 核函数

支持向量机通过某非线性变换 $\phi(x)$ ，将输入空间映射到高维特征空间。特征空间的维数可能非常高。如果支持向量机的求解只用到内积运算，而在低维输入空间又存在某个函数 $K(x, x')$ ，它恰好等于在高维空间中这个内积，即 $K(x, x') = \langle \phi(x), \phi(x') \rangle$ 。那么支持向量机就不用计算复杂的非线性变换，而由这个函数 $K(x, x')$ 直接得到非线性变换的内积，使大大简化了计算。这样的函数 $K(x, x')$ 称为核函数。