Kernel Methods

FrankZhou-jun*

2019 年 10 月 14 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 Eiganvalue Decomposition(特征值分解) **6**

4 Sigualr Value Decomposition(奇异值分解) **6**

5 SVD 用于 PCA 降维 **6**

6 核函数 **7**

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 $\boldsymbol{P_1}(x_1, y_1, z_1)$ 和垂直于该平面的法线 $\vec{\boldsymbol{n}}$，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

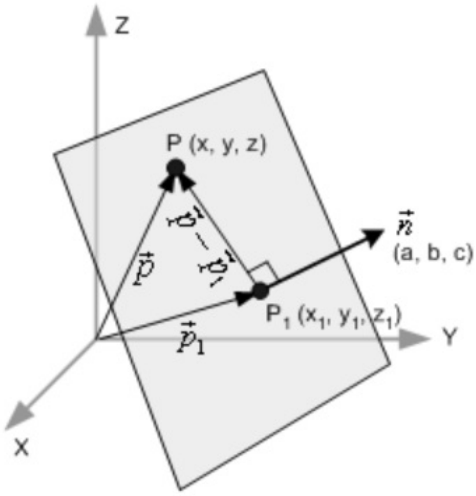


图 1: 点法式

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

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

法向量 $\vec{\boldsymbol{n}}$ 与点 $\boldsymbol{P_1}(\boldsymbol{x_1}, \boldsymbol{y_1}, \boldsymbol{z_1})$ 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 $\boldsymbol{P}(x, y, z)$ 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

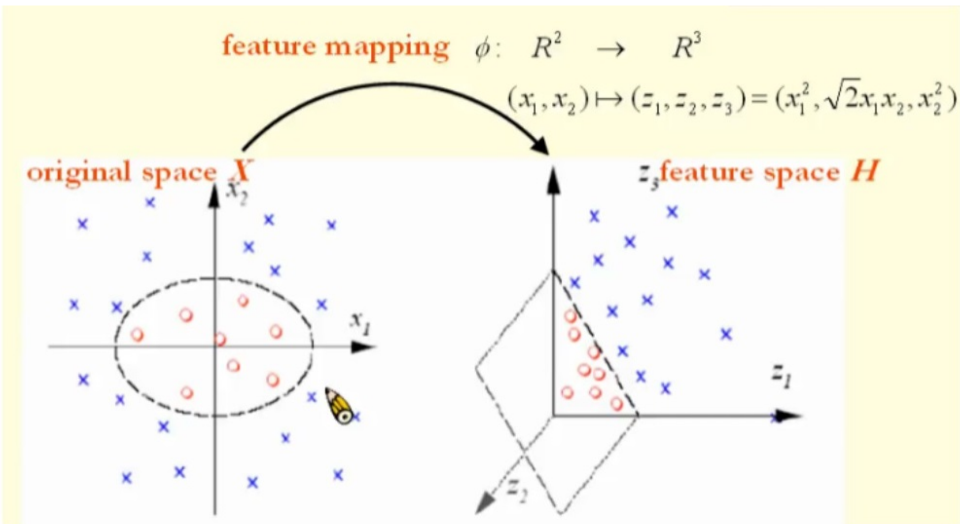


图 2: 超平面

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，可能需要弯弯拐拐的才行。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

$$y=-\frac{a}{b}x-\frac{c}{b}z-\frac{d}{b} \tag{3}$$

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：

$$\boldsymbol{W^T x}+d=0 \tag{4}$$

其中 $\boldsymbol{W, x}$ 的维度都为 n

$$\boldsymbol{w} \quad =(w_1,w_2,\ldots,w_n)^T \tag{5}$$
$$\boldsymbol{x} \quad =(x_1,x_2,\ldots,x_n)^T \tag{6}$$

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

$$\vec{\boldsymbol{n}} \cdot (\boldsymbol{P}-\boldsymbol{P_1})=0 \tag{7}$$

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 $\boldsymbol{P_2}$ 到超平面的距离 D 等于超平面上的点 $\boldsymbol{P_1}$ 指向 $\boldsymbol{P_2}$ 构成的向量 $\overrightarrow{\boldsymbol{P_1P_2}}$ 在该平面法向量 $\vec{n}$ 上投影的距离长度。参考图3

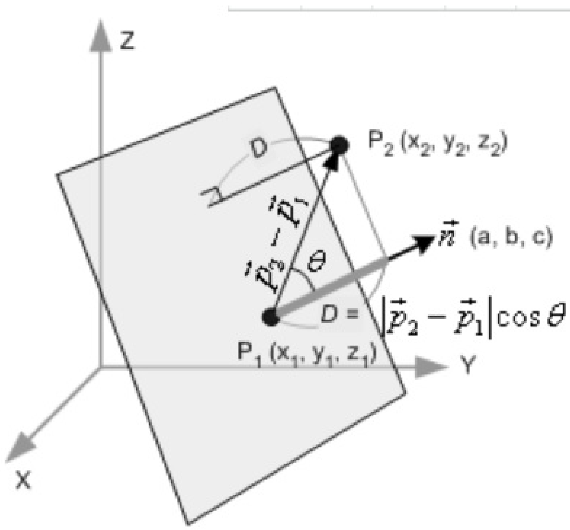


图 3: 点到平面的距离

距离 D 的大小表示为

$$D=|\overrightarrow{\boldsymbol{P_1P_2}}|\cos\theta \tag{8}$$

[2]https://www.quora.com/What-are-kernels-in-machine-learning-and-SVM-and-why-do-we-need-them/answer/Lili-Jiang?srid=oOgT

[3]百度

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

$$\cos\theta = \frac{\vec{\boldsymbol{n}} \cdot \overrightarrow{\boldsymbol{P_1P_2}}}{|\vec{\boldsymbol{n}}||\overrightarrow{\boldsymbol{P_1P_2}}|} \tag{9}$$

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

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

$$D = \frac{\boldsymbol{W^T x} + d}{|\vec{\boldsymbol{n}}|} \tag{11}$$

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

2 特征值与特征向量

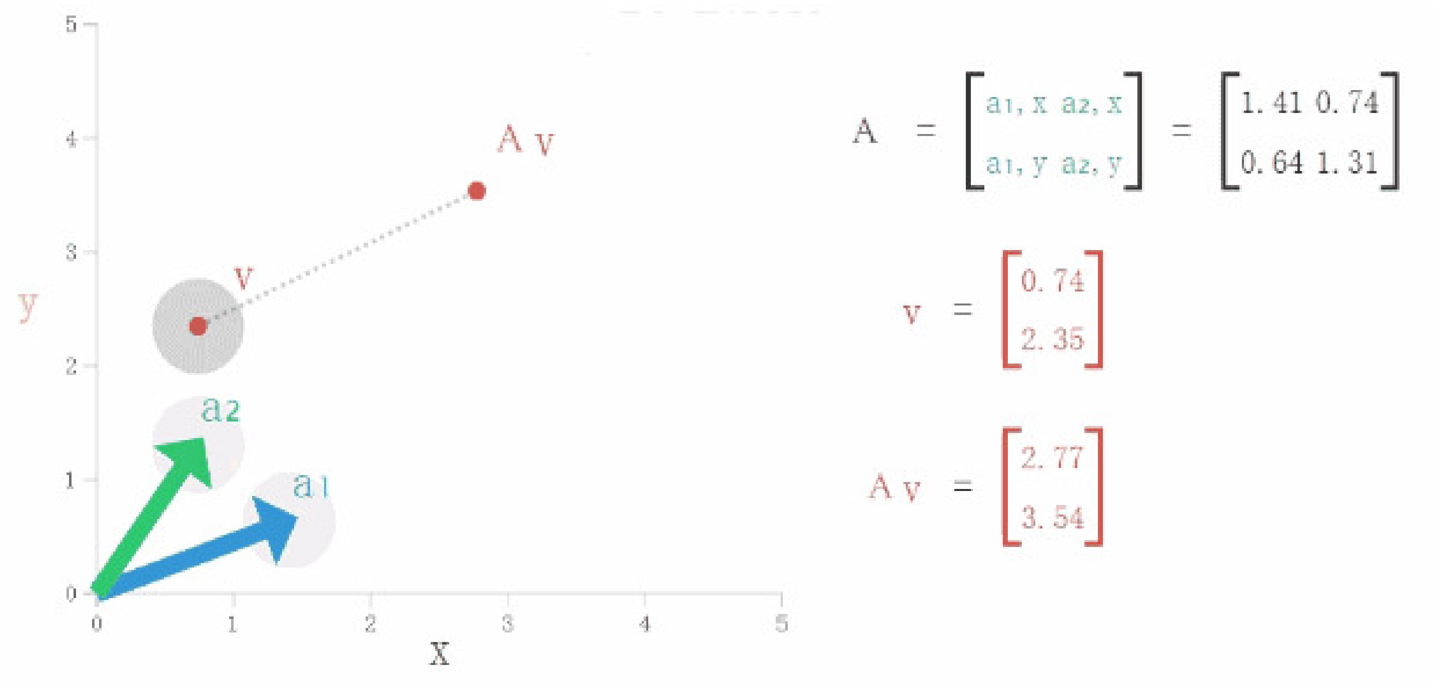
先来讲讲定义：设 $\boldsymbol{A}$ 为 n 阶方阵, 若存在 n 阶非零向量 $\vec{\boldsymbol{x}}$ 使得

$$\boldsymbol{A}\vec{\boldsymbol{x}} = \lambda\vec{\boldsymbol{x}} \tag{13}$$

则称 λ 为矩阵 $\boldsymbol{A}$ 的特征值，$\vec{\boldsymbol{x}}$ 为对应特征值下矩阵 $\boldsymbol{A}$ 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.

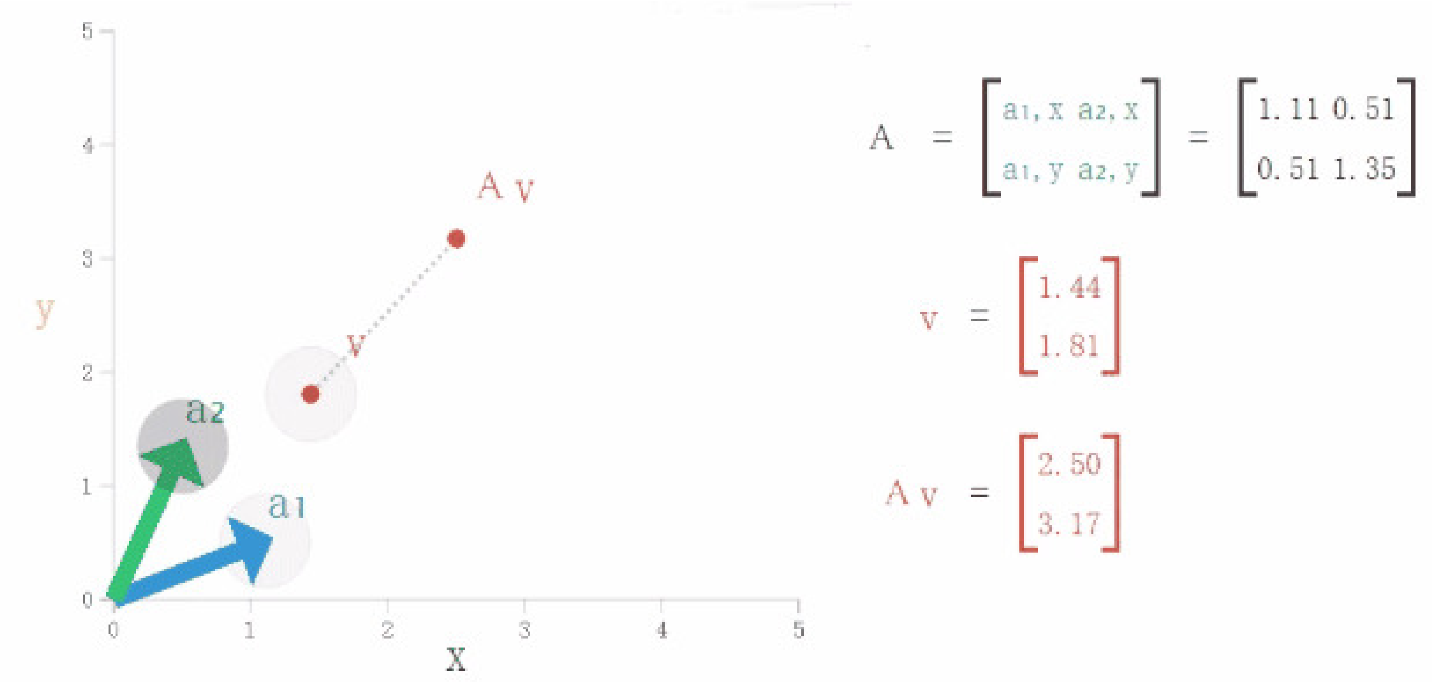
先上两张图

假设矩阵 $\boldsymbol{A}$ 不变，$\vec{\boldsymbol{x}}$ 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 $\boldsymbol{A} = \begin{vmatrix} 1 & -2 \\ 1 & 4 \end{vmatrix}$ 的特征值和特征向量。

$$|\lambda \boldsymbol{E} - \boldsymbol{A}|\vec{\boldsymbol{x}} = 0 \tag{14}$$

方程 $|\lambda \boldsymbol{E} - \boldsymbol{A}| = 0$ 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 $|\lambda \boldsymbol{E} - \boldsymbol{A}|\vec{\boldsymbol{x}} = 0$ 非零解。

先求 λ

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

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**

$$\begin{gathered} \begin{vmatrix} 1 & 2 \\ -1 & -2 \end{vmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \\ x1 = 2x_2 \end{gathered} \tag{16}$$

即，求得**基础解析**

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

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**

$$\begin{gathered} \begin{vmatrix} 2 & 2 \\ -1 & -1 \end{vmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \\ x1 = -x_2 \end{gathered} \tag{18}$$

即，求得基础解析

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

3 Eiganvalue Decomposition(特征值分解)

今天了解了一下矩阵特征值分解，若 $\mathbf{A}$ 为 $m \times m$ 维度的实对称矩阵 ($\mathbf{A} = \mathbf{A^T}$), 则可以被分解为一下形势：

$$\begin{aligned} \mathbf{A} =& \mathbf{Q}\Sigma\mathbf{Q^T} \\ =& \mathbf{Q}\begin{bmatrix} \lambda_1 & \cdots & \cdots & \cdots \\ \cdots & \lambda_2 & \cdots & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & \cdots & \cdots & \lambda_i \end{bmatrix}\mathbf{Q^T} \end{aligned} \tag{20}$$

其中 $\mathbf{QQ^T}$ 为标准正交矩阵，即 $\mathbf{QQ^T} = \mathbf{I}$，$\Sigma$ 为对角矩阵，这三个矩阵的维度均为 $m \times m$，λ_i 为特征值，$\vec{q_i}$ 为 $\mathbf{Q}$ 每一列向量，每一个特征值就对应一个特征向量。

$$\mathbf{A}\vec{\boldsymbol{q_i}} = \Sigma\vec{\boldsymbol{q_i}} \tag{21}$$

可以看到特征值分解需要实对称矩阵，而实际情况中矩阵通常是 $m \times n$, 分解调节比较苛刻，所以出现了 Sigular value decomposition(奇异值分解)

4 Sigualr Value Decomposition(奇异值分解)

假设矩阵 $\mathbf{A}$ 的维度为 $m \times n$，存在酉矩阵[5] $\mathbf{U}, \mathbf{V}$ 使得

$$\begin{aligned} \mathbf{A} =& \mathbf{U}\Sigma\mathbf{V^T} \\ =& \mathbf{U}\begin{bmatrix} \lambda_1 & \cdots & \cdots & \cdots \\ \cdots & \lambda_2 & \cdots & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & \cdots & \cdots & \lambda_i \end{bmatrix}\mathbf{V^T} \end{aligned} \tag{22}$$

其中 $\mathbf{UU^T} = \mathbf{E}, \mathbf{VV^T} = \mathbf{E}$,$\Sigma$ 为对角矩阵，这里 λ_i 称为奇异值，$\mathbf{U}$ 为左奇异矩阵，$\mathbf{V^T}$ 为右奇异矩阵。其中 $\mathbf{U}$ 的维度为 $m \times m$,Σ 的维度为 $m \times n$,$\mathbf{V^T}$ 的维度为 $n \times n$
我们可以选择有限个奇异值或特征值重组矩阵，进行降噪，降维... 等操作。

5 SVD 用于 PCA 降维

先找到需要找到样本协方差矩阵 $\mathbf{XX^T}$ 的最大的 d 个特征向量，然后用这最大的 d 个特征向量张成的矩阵来做低维投影降维。

$$X'_{m\times n} \approx U_{m\times d} * \Sigma_{d\times d} * V^T_{d*n} \tag{23}$$

选用一定的奇异值会导致矩阵与原矩阵出现偏值，可用选用的奇异值与总的奇异值占比来确定。
假设降到 k 维

$$X'_{m\times k} \approx U_{m\times d} * \Sigma_{d\times d} * V^T_{d*k} \tag{24}$$

[5]其中酉矩阵的数域为复数

6 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 $K(x,x\ ) =< \phi(x), \phi(x\ ) >$ 。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.