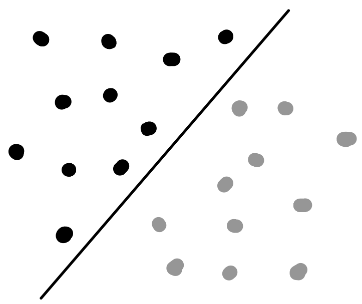
支撑向量机

二分类模型，基本模型是定义在特征空间上的间隔最大的分类器，学习策略是间隔最大化

一、线性分类器



用一条直线将黑点和白点分开，可以有无数条

令黑色的点标签为-1，白色的点标签为1

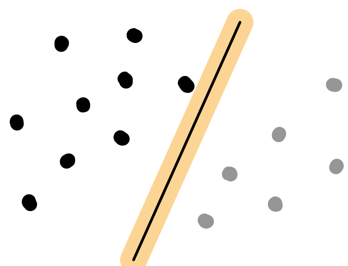
直线为 $f(\vec{x}) = \vec{w}^T\vec{x} + b$

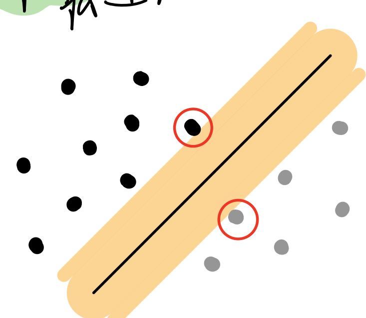
$\vec{x}$的维度为n时，$f(\vec{x})$表示n-1维超平面

符号函数 $\mathrm{sgn}(x) = \begin{cases} -1, & x<0 \\ 0, & x=0 \\ 1, & x>0 \end{cases}$

如何确定一条最优的直线？

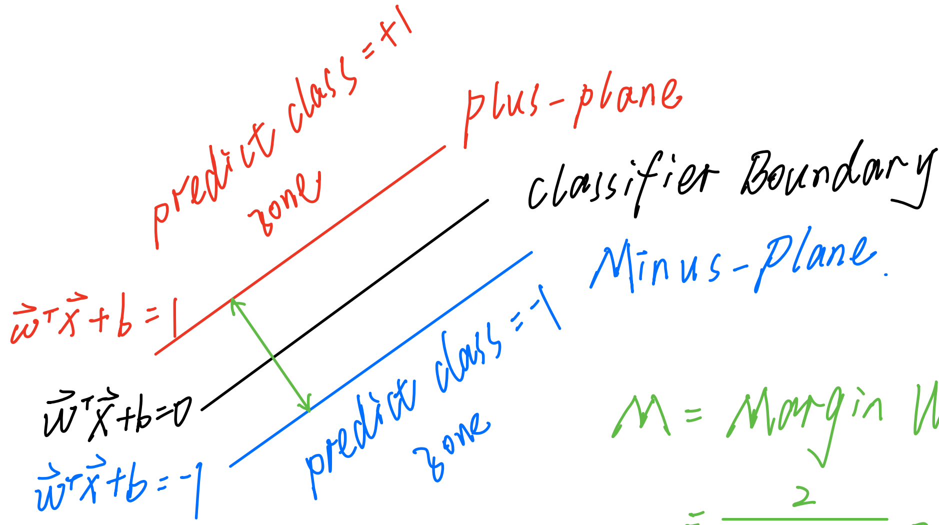
让这条直线到给定样本中最近的点最远。





直观上来讲，分隔的间隙越大越好。

上图中被圈出来的点即为支撑向量 $\boxed{\text{个数可以大于2}}$



M = Margin Width

$$= \frac{2}{\sqrt{\vec{w}^T\cdot\vec{w}}} = \frac{2}{\|\vec{w}\|}$$

classifier boundary 是 $f(x)$

plus plane 和 Minus plane 是支撑向量所在的平面

之间的距离便是要最大化的分隔间隙

支撑向量的表达式为 $y(\vec{w}^T\vec{x}+b)$ $y\in\{-1,1\}$

目标函数：$\max M \to \max \frac{2}{\|\vec{w}\|} \to \min \frac{1}{2}\|\vec{w}\|^2$

原问题：

$$\min \frac{1}{2}\|\vec{w}\|^2 \quad s.t.\ y_i(\vec{w}^T x_i+b)\geq 1,\ i=1,2,\dots,n \quad (1)$$

其实是一个带约束的二次规划，是凸问题，可得最优解

二 转化为对偶问题

上述优化问题可用拉格朗日乘子法去解：

对应的拉格朗日目标函数为：

$$L(\vec{w},b,\vec{\alpha})=\frac{1}{2}\|\vec{w}\|^2-\sum_{i=1}^{n}\alpha_i(y_i(\vec{w}^T\vec{x}_i+b)-1) \quad (2)$$

根据拉格朗日对偶性，得到对偶问题的表述式：

$$L(\vec{w},b,\vec{\alpha})=\frac{1}{2}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \vec{x}_i\cdot\vec{x}_j \quad (3)$$

加上限制条件后，

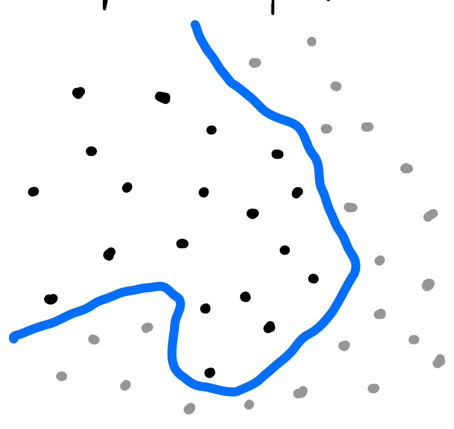
$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \vec{x}_i \vec{x}_j$$

$$s.t.\ \alpha_i\geq 0,\ i=1,2,\cdots,n,\ \sum_{i=1}^{n}\alpha_i y_i=0. \quad (4)$$

上式即为需要优化的表述式

三 线性不可分的情况

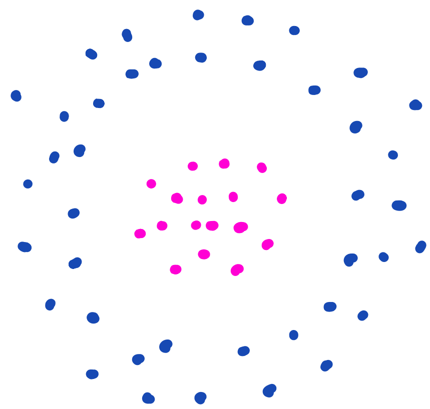
线性可分的假设非常有局限性



四. 核函数

使用某些非线性方法，可以得到将两个类别完美划分的曲线，比如核函数，

让空间从线性空间变成一个更高维的空间，在高维的线性空间下，使用超平面进行划分



对偶问题表达式，式(4)

$$\max_\alpha \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j y_i y_j \vec{x_i}\vec{x_j}$$

$$s.t.\ \alpha_i \geq 0,\ i=1,2,\cdots,n,\ \sum_{i=1}^n \alpha_i y_i = 0.$$

将红色圈中的部分进行改造

令：$\vec{x_i}\cdot\vec{x_j} = K(\vec{x_i}\cdot\vec{x_j})$ (5)

这个式子就是将线性空间转换到高维空间

映射到高维空间中的方法有很多种，如：

多项式核：$K(\vec{x_i}, \vec{x_j}) = (\vec{x_i} \cdot \vec{x_j} + 1)^d$

高斯核：$K(\vec{x_i}, \vec{x_j}) = \exp(-\frac{\|\vec{x_i} - \vec{x_j}\|^2}{2\sigma^2})$

高斯核甚至是将原始空间映射为无穷维空间。

核函数不会增加过多计算量

五．其他问题

1) 如何进行多分类。

主要有两种方式：1种是 1 v.s. (N-1)，2种是 1 v.s 1

第1种方法训练 N 个模型，属于第i类或其他类

第2种方法训练 $\frac{N(N-1)}{2}$ 个模型，属于第i类或第j类

2) SVM 会 overfitting 吗？

SVM是一种不太容易 overfitting 的方法。

3) 优缺点

优点：

1. 本身是凸优化问题，可得到全局最优解

2. 适用于线性问题和非线性问题
3. 高维样本空间的数据也能使用SVM,
因为复杂度只取决于支持向量而非所有数据
4. 理论基础比较完善.

缺点:

1. 二次规划问题涉及m阶矩阵的计算
m为训练集的个数.所以不适于超大数据集
(SMO算法可缓解该问题)
2. 只适用于二分类问题,但可以解决.