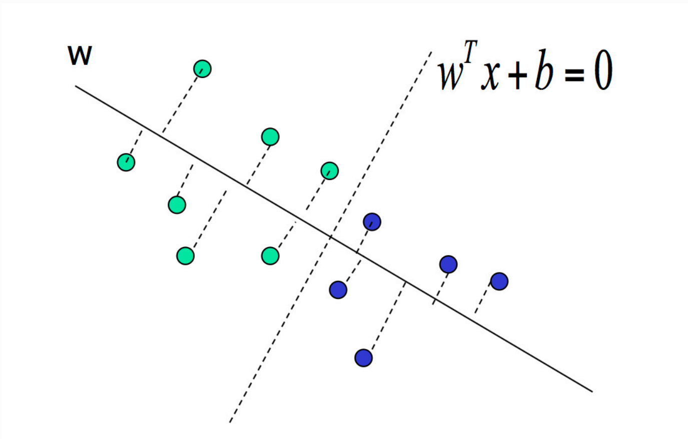
线性判别函数

线性判别函数(两类)

$$g(x) = w^T x + b = \begin{cases} > 0 & w_1 \\ < 0 & w_2 \end{cases}$$

其中，$g(x) = 0$ 就是分类面方程

直观展示	解释
	可以看出我们需要做的就是选择w, 而选择w本质上就是寻找一个最佳的投影方向. 如左图，数据投影到法向量之后就是一维数据的分类问题

Fisher 线性判别

Fisher判别的基本思想

- 直观描述
  - 投影后的数据满足
    - 两类之间的距离尽可能远
    - 每类自身尽可能紧凑
- 数学描述
  - 考虑投影后数据的均值和方差
    - 两类数据的均值
    - 两类数据的协方差

理论推导

`以两类为例子`

记原始数据的均值为m_1, m_2，协方差为Σ_1, Σ_2

记原始数据投影后数据的均值为μ_1, μ_2，协方差为σ_1, σ_2

其中易得

$$\mu_i = w^T m_i \text{，此处}w\text{为单位向量}$$
$$\sigma_i^2 = E(w^T x - E(w^T x))^2 = w^T E(x - E(x))^2 w = w^T \Sigma_i w$$

> 注: $x^T y = (x, y)$，而内积可以看作一个向量在另一个向量上作投影，直观地看 $(x, y) = \| x \| \| y \| \, cos(< x, y >)$

Fisher准则函数：

$$J_F(w) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$
$$w_{opt} = arg\,max\,J_F(w)$$

由准则函数进一步推导可得

$$J_F(w) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} = \frac{(w^T m_1 - wTm_2)^2}{w^T \Sigma_1 w + w^T \Sigma_2 w} = \frac{w^T (m_1 - m_2)(m_1 - m_2)^T w}{w^T (\Sigma_1 + \Sigma_2) w}$$

据此可以定义

- 类间离散度矩阵

$$S_b = (m_1 - m_2)(m_1 - m_2)^T$$

- 类内总离散度矩阵

$$S_w = \Sigma_1 + \Sigma_2 \quad \text{，} S_w\text{正定}$$

  - 若$w^T S_w w = 0$意味着所有的数据都被投影到了一点

$J_F(w)$有上界，所以最佳投影方向一定存在

$$J_F(w) = \frac{w^T S_b w}{w^T S_w w}$$
$$\Downarrow$$
$$J_F(w) \le \frac{\lambda(S_b)_{max}}{\lambda(S_w)_{min}}$$

最优化问题构建

一定存在最优的w使得

$$w^T S_w w = 1$$

故我们的目标函数可以写为

$$\begin{aligned} max\ w^T S_b w \\ s.t.\ \ w^T S_w w = 1 \end{aligned}$$

拉格朗日乘子法转变如下

$$L(w,\lambda) = w^T S_b w - \lambda(w^T S_w w - 1)$$

求导便得到

$$\frac{\partial L(w,\lambda)}{\partial w} = S_b w - \lambda S_w w = 0$$

> 此处因为S_w正定必可逆，如果我们等式两边同时乘S_w^{-1},可得

> $$S_w^{-1} S_b w = \lambda w \quad \text{可见}w\text{是矩阵}S_w^{-1}S_b\text{的特征向量}$$

继续之前公式可得

$$(m_1 - m_2)(m_1 - m_2)^T w_{opt} = \lambda S_w w_{opt}$$

其中$(m_1 - m_2)^T w_{opt}$是一个数，同时参数λ也是一个数，我们求解w时只关注它的方向，所以可以把两边的数字全部去掉，于是得到

$$\begin{aligned} (m_1 - m_2) = S_w w_{opt} \\ \Downarrow \\ w_{opt} = S_w^{-1}(m_1 - m_2) = (\Sigma_1 + \Sigma_2)^{-1}(m_1 - m_2) \end{aligned}$$

参数b的确定方式就很简单了

$$b = \frac{n_1\mu_2 + n_2\mu_2}{n_1 + n_2}, \text{其中}n_1, n_2\text{是两类样本的个数}$$

最小平方误差准则

线性回归模型的例子

$$f(x) = \beta_0 + \sum_{i=1}^{n} x_i \beta_i$$

将 1 放入X 即增广矩阵，可写成

$$\hat{Y} = X\beta$$

基于Least square 得到

$$min\ J(\beta) = \| Y - X\beta \|^2$$

求导可得

$$\frac{\partial J}{\partial \beta} = -2X^T(Y - X\beta_{opt}) = 0$$

于是有

$$Y = X\beta_{opt}$$

避免X不可逆，两边同乘X^T

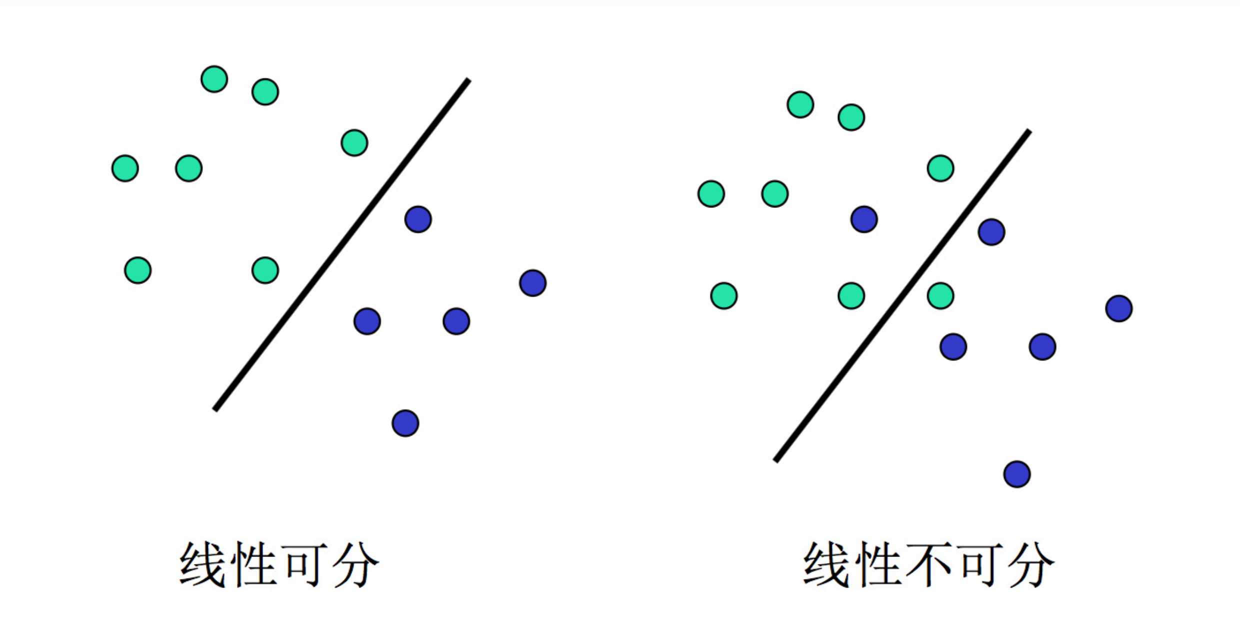
$$X^TY = X^TX\beta_{opt}$$

所以可以得到

$$\beta_{opt} = (X^TX)^{-1}X^TY = \frac{(X, Y)}{(X, X)}$$

最小错分样本数准则

有时，数据并不是线性可分的，比如



这样的情况下，我们最直接的想法就是考虑最小错分量