线性分类器

两步贝叶斯决策：首先根据样本进行概率密度估计，在根据估计的概率密度求分界面。

概率密度函数做决策，最终也是也是依靠判别函数。p29,理解知识点的关联关系。

判别式模型:直接从数据中估计判别函数的参数，而不必估计概率密度函数。

基于样本设计分类器需要三个要素：

1.判别函数（分类器）的类型

2.分类器设计的目标或准则

3.设计算法利用数据搜索到最优的参数（优化）

判别函数类、准则、优化算法

4.2 判别函数的基本概念

TransR

$$x = x_p + r\frac{w}{\|w\|}$$

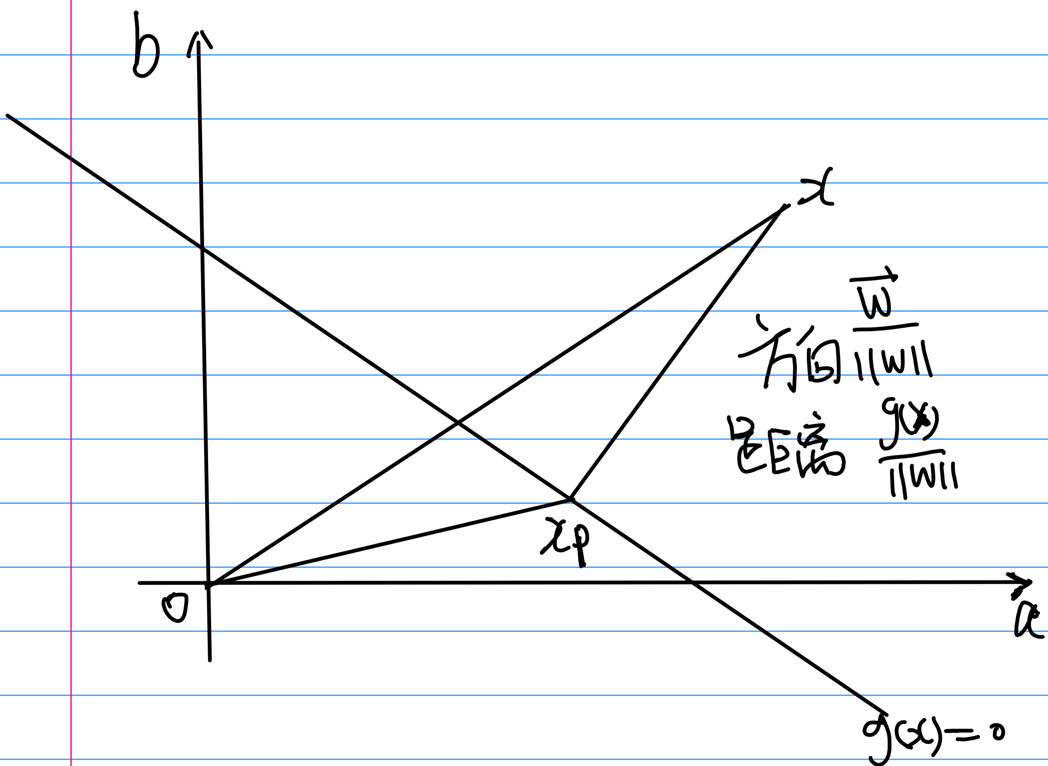
$$g(x) = w^T x + w_0 = w^T\left(x_p + r\frac{w}{\|w\|}\right) + w_0$$

$$= w^T x_p + w_0 + r\frac{w^T w}{\|w\|}$$

$$= r\frac{w^T w}{\|w\|} = r\frac{\|w\|^2}{\|w\|} = r\|w\|$$

$$r = \frac{g(x)}{\|w\|}$$

r是x到H的垂直距离 x_p是射影向量



$x=(a,b)$

二维向量

超平面的方向由 w确定，它的位置由w_0确定。

判别函数g(x)正比于x点到超平面的代数距离，带正负号，也就是r，类似于代数余子式的说法。

4.3 Fisher线性判别分析

主要解决两类的线性判别问题，Fisher线性判别的思想是选择投影方向，使投影后两类相隔尽可能远，而同时每一类内部的样本又尽可能聚集。

应该找哪个方向去投影最好呢？

$y_i = w^T x_i$ ，w就是投影方向

x_i样本，y_i是新样本

Fisher判别准则 （要素2） 要素1就是y=w^T x+w_0

$$\max_w J_F(w) = \frac{w^T S_b w}{w^T S_w w}$$

倒着看书

倒背如流有道理

广义 Rayleigh 商.

S_b：类间散度 between-class

S_w：类内散度 within-class

$$\max J_F(w) = \frac{\tilde{S}_b}{\tilde{S}_w} = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

$$\left.\begin{array}{l} w^T S_b w = \tilde{S}_b \\ w^T S_w w = \tilde{S}_w \end{array}\right\} \text{why?}$$

倒着看书，就会问why

$$\tilde{S}_b = (\tilde{m}_1 - \tilde{m}_2)^2 = (w^T m_1 - w^T m_2)^2$$

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

$$= w^T S_b w$$

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

$$m_i = \frac{1}{N_i} \sum_{x_j \in X_i} x_j$$

$$\tilde{S}_w = \tilde{S}_1^2 + \tilde{S}_2^2$$

$$= \sum_{x_j \in X_1} (w^T x_j - w^T m_1)^2 + \sum_{x_j \in X_2} (w^T x_j - w^T m_2)^2$$

$$= \sum_{x_j \in X_1} w^T (x_j - m_1)(x_j - m_1)^T w + \sum_{x_j \in X_2} w^T (x_j - m_2)(x_j - m_2)^T w$$

$$= w^T S_1 w + w^T S_2 w$$

$$= w^T S_w w$$

$$S_w = S_1 + S_2$$

优化：（要素3）

最优投影方向 $w^* = S_w^{-1}(m_1 - m_2)$

$\max \ w^T S_b w$

$\text{s.t.} \ w^T S_w w = c \neq 0$ subject to

固定分母

Lagrange函数的无约束极值问题，

优化过程按步就班去做

$$g(x) = w^T x + w_0 \gtrless 0 \text{ 则 } x \in \begin{cases} w_1 \\ w_2 \end{cases}$$

当样本是正态分布且类协方差矩阵相同，最优贝叶斯分类器如下

$$W = \Sigma^{-1}(\mu_1 - \mu_2)$$

$$W_0 = -\frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) + \ln\frac{P(w_1)}{P(w_2)}$$

因此在这种情况下，Fisher线性判别的方向就是最优贝叶斯方向

$$W_0 = -\frac{1}{2}(m_1 + m_2)^T S_w^{-1}(m_1 - m_2) - \ln\frac{P(w_2)}{P(w_1)}$$

$w^* = S_w^{-1}(m_1 - m_2)$ 投影方向

$\tilde{m}_1 = m_1^T \cdot w^*$ m_1沿w^*投影

$W_0 = -\frac{1}{2}(\tilde{m}_1 + \tilde{m}_2)$ 当 $P(w_1) = P(w_2)$

$= -\tilde{m}$

$\tilde{m} = \frac{\tilde{m}_1 \cdot n_1 + \tilde{m}_2 \cdot n_2}{n_1 + n_2}$ 当$n_1 = n_2$ $= \frac{1}{2}\tilde{m}_1 + \frac{1}{2}\tilde{m}_2$

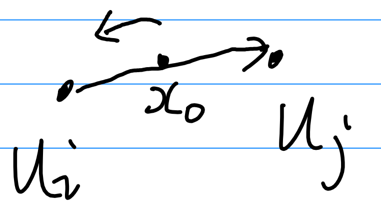
总体平均值是各类平均值的加权组合

先验概率大的要占据更大的决策空间，因此当先验概率不相等，决策面向先验概率小的那边偏

(2-80)，(2-88)

x_0是决策面上的点，$W(x_0-x_0)=0$成立

$$x_0=\frac{1}{2}(\mu_i+\mu_j)-k\cdot\ln\frac{P(w_i)}{P(w_j)}(\mu_i-\mu_j)$$



若 $P(w_i)>P(w_j)$

则 $\vec{x}_0'=\vec{x}_0-\overrightarrow{(\mu_i-\mu_j)}$

$$k\cdot\ln\frac{P(w_i)}{P(w_j)}>0$$

(4-2) $\vec{W}$指向R_1为正例

若 $\ln\frac{P(w_2)}{P(w_1)}>0$ 则 $W^T\cdot x>0$才行 故x落入正例更深才行

4.4感知器