二次判别函数

$$
\begin{aligned}
g(\boldsymbol{x}) &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x} + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + w_0 \\
&= \sum_{k=1}^{d} w_{kk} x_k^2 + 2\sum_{j=1}^{d-1}\sum_{k=j+1}^{d} w_{jk} x_j x_k + \sum_{j=1}^{d} w_j x_j + w_0
\end{aligned}
$$

其中,$\boldsymbol{W}$ 是 $d\times d$ 维实对称矩阵,$\boldsymbol{w}$ 为 d 维向量。

不难看出,这个判别函数中包含0.5d(d+3)+1个参数,因此如果像线性判别函数那样，直接根据一定的规则从数据去学习这些参数,计算起来会非常复杂,而且在样本数不足够多时估计如此多的参数,结果的可靠性和推广能力很难保证。

为什么有这些参数？
因为实对称矩阵参数的个数是，0.5d(1+d), w有d个参数， 最后有1个w0,
总计0.5d(d+3)+1

实际中,人们在应用二次判别函数时,往往采用参数化的方法来估计二次判别函数，知道形式，不知道参数。比如，人们往往假定每一类数据都是正态分布，每一类的判别函数如下

$$g_i(\boldsymbol{x}) = K_i^2 - (\boldsymbol{x}-\boldsymbol{m}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{m}_i)$$

K2是一个阈值项,它受协方差矩阵和先验概率的影响。判别函数就是样本到均值的Mahalanobis 距离的平方与固定阈值的比较,样本的均值和方差可以用下面的估计

$$\hat{\boldsymbol{m}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} \boldsymbol{x}_j$$

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{N_i - 1}\sum_{j=1}^{N_i} (\boldsymbol{x}_j - \hat{\boldsymbol{m}}_i)(\boldsymbol{x}_j - \hat{\boldsymbol{m}}_i)^{\mathrm{T}}$$

1. 当两类都近似正态分布时，两类样本的分布各自比较集中，决策面方程为

$$\text{若 } g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) \gtrless 0, \text{则 } \boldsymbol{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

2.两类中w1一类分布比较成团(近似正态分布),另一类w2则比较均匀地分布在第一类附近,这种情况下只要对第一类求解其二次判别函数即可,即

$$g(\boldsymbol{x}) = K^2 - (\boldsymbol{x} - \hat{\boldsymbol{m}}_1)^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_1^{-1}(\boldsymbol{x} - \hat{\boldsymbol{m}}_1)$$

$$\text{若 } g(\boldsymbol{x}) \gtrless 0, \text{则 } \boldsymbol{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

同样,可以用 K^2 来调整决策的偏向。直观解释是,当样本到 ω_1 类均值的 Mahalanobis 距离的平方小于 K^2 时则决策为 ω_1 类,否则决策为 ω_2 类。