4.4 感知器

1.判别函数

增广样本向量

$$g(y) = a^T y \;;\; y \text{ is } [1, x_1, \cdots x_d]^T$$

$$a \text{ is } [w_0, w_1, \cdots w_d]^T$$

2.准则

$$J_p(a) = \sum_{a^T y'_k < 0} (-a^T y'_k)$$

3.优化算法

梯度下降法

$$g(y) \gtrless 0 \begin{cases} y \in w_1 \\ y \in w_2 \end{cases}$$

$$y_i' = \begin{cases} y_i, & 若\ y_i \in \omega_1 \\ -y_i, & 若\ y_i \in \omega_2 \end{cases}$$

规范化增广样本向量

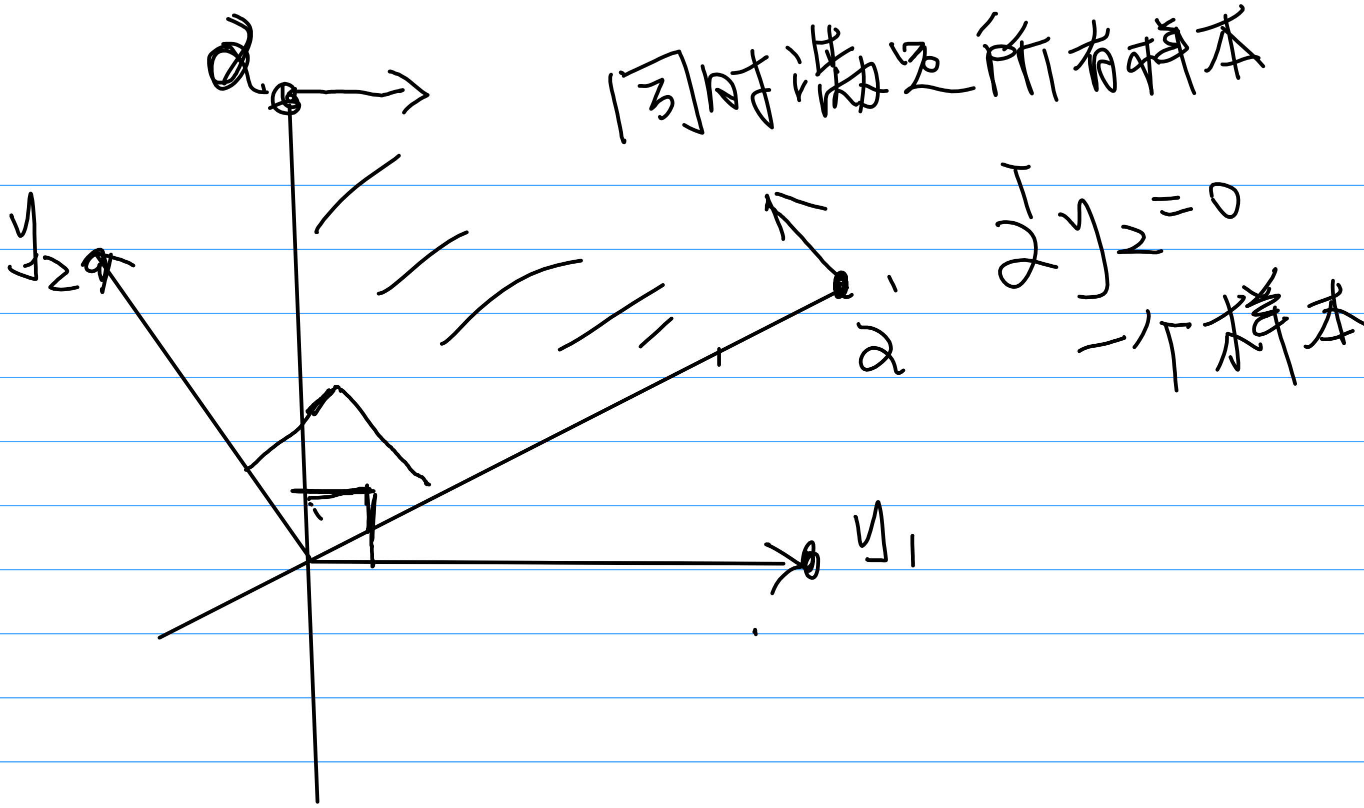
样本可分性条件

$$a^T y_i' > 0$$

此方法适应于线性可分的样本，即在样本特征空间中，至少存在一个线性分类面能够把两类样本没有错误的分开

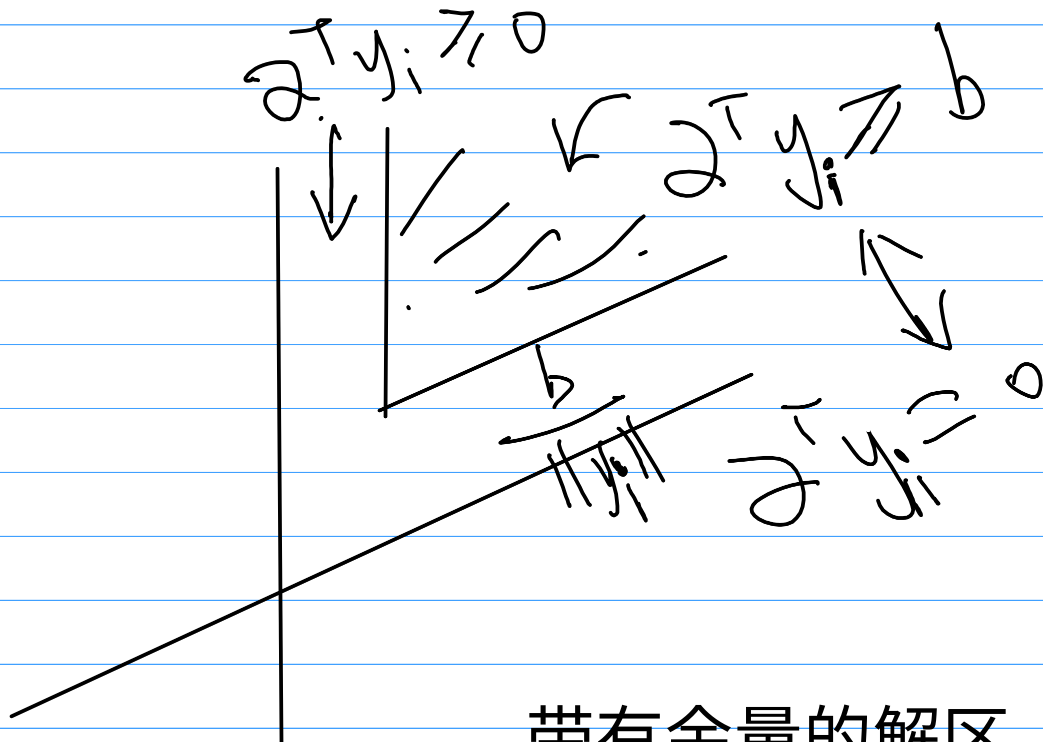
$$a^T y_i > 0, \quad i = 1, 2, \cdots N$$

a^* 为一个解向量



什么是解向量和解区？

如何选取更可靠的解向量？



带有余量的解区

如何优化准则函数？

$$J_p(a^*) = \min J_p(a) = 0$$

即 a^* 将所有样本分对

梯度下降法迭代求解

$$a(t+1) = a(t) - \rho_t \nabla J_p(a)$$

$$\nabla J_p(a) = \frac{\partial J_p(a)}{\partial a} = \sum_{a^T y_k \le 0} (-y_k)$$

$$a(t+1) = a(t) + \rho_t \sum_{a^T y_k \le 0} y_k$$

一次将所有错误样本都进行修正的做法并不是效率最高的，更常用的是一次只修正一个样本的固定增量法。

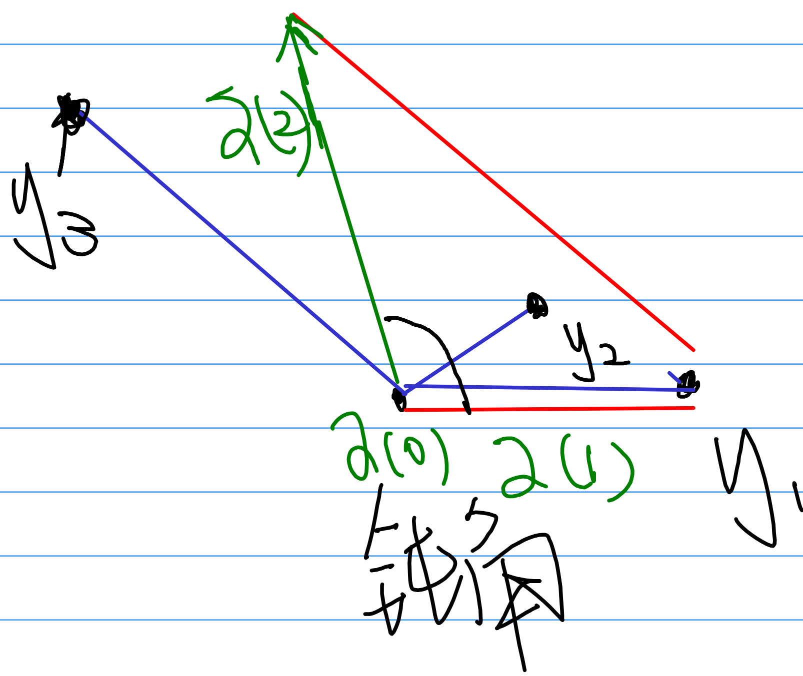
（1）任意选择初始的权向量 $a(0)$，置 $t=0$

（2）考察样本 y_j，若 $a(t)^T y_j \le 0$ 则 $a(t+1)=a(t)+y_j$，否则继续，$\boxed{\rho_t=1}$

(3) 考察另一个样本，重复(2)，直至所有样本都有 $a(t)^T y_j > 0$，

即 $J_p(a)=0$ （$a(t)^T y_j \le b$）

错分类余量



对于线性可分的样本集迭代下去会收敛，
对于线性不可分的样本集则不会收敛

$$\rho_t = \frac{|a(k)^T y_j|}{\|y_j\|^2}$$

对 ρ_t 改进

4.5最小平方误差判别准则

线性不可分情况下，方程个数大于未知数，属于矛盾方程组，求解方程组的最小平方误差解

$a^T y_i > 0$，不等式组不能同时满足

希望求解一个a^*使得被错分的样本尽可能少

$a^T y_i = b_i > 0$

$Ya = b$

$$Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_N^T \end{bmatrix} = \begin{bmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{bmatrix}$$

$$b = [b_1, \cdots b_N]^T$$

$N > d$, dimension

方程组误差 $e = Ya - b$

$$\min J_s(a) = \|Ya - b\|^2$$

$$= \sum_{i=1}^{N} (a^T y_i - b)^2$$

伪逆法求解

$$\nabla J_s(a) = 2Y^T(Ya-b) = 0$$

$$a^* = (Y^TY)^{-1}Y^Tb.$$

$(Y^TY)^{-1}Y^T$ 是 Y 的伪逆

梯度下降法求解

（1）任意选择初始的权向量 $a(0)$，置 $t=0$

（2）按照梯度下降的方向迭代更新权向量

$$a(t+1) = a(t) - \rho_t Y^T(Ya-b)$$

$\nabla J_s(a) \le \xi$ 或

$\|a(t+1) - a(t)\| \le \xi$

则停止.

$$a(t+1) = a(t) + \rho_t (b_k - a(t)^T y_k) y_k$$

其中 $a(t)^T y_k \ne b$

Widrow-Hoff算法

后面讨论了关于选择b的问题，选不同的b，MSE的解等价于Fisher线性判别的解或者贝叶斯判别函数的最小平方误差逼近。