$$y = x^T A x + b^T x + c$$

$$P(A|x) = \frac{P(x|A)\,P(A)}{P(x)} = \frac{P(x|A)\cdot P(A)}{P(x|A)P(A) + P(x|B)P(B)}$$

$$= \frac{1}{1 + \frac{P(x|B)P(B)}{P(x|A)P(A)}}$$

取 $\ln \frac{P(x|B)\,P(B)}{P(x|A)\,P(A)} = \alpha$

$$P(A|x) = \frac{1}{1 + e^{\alpha}}.$$

α 决定 $P(A|x)$

$$\alpha = \ln \frac{P(x|B)\,P(B)}{P(x|A)\,P(A)}$$

$$= \ln \frac{P(x|B)}{P(x|A)} + \ln \frac{P(B)}{P(A)}$$

Gauss: $\frac{1}{(2\pi)^{\frac{D}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$

$$= \ln \frac{\cancel{\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{\Sigma}}} \exp\left(-\frac{1}{2}(x-\mu_B)^T \Sigma_B^{-1} (x-\mu_B)\right)}{\cancel{\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{\Sigma}}} \exp\left(-\frac{1}{2}(x-\mu_A)^T \Sigma_A^{-1} (x-\mu_A)\right)} + \ln \frac{P(B)}{P(A)}$$

$$= \left(-\frac{1}{2}(x-\mu_B)^T \Sigma_B^{-1} (x-\mu_B)\right) + \frac{1}{2}\left((x-\mu_A)^T \Sigma_A^{-1} (x-\mu_A)\right) + \cancel{\ln \frac{P(B)}{P(A)}}$$

$$= \left(-\frac{1}{2}\right)\left(x^T \Sigma_B^{-1} \cdot x\right) + \frac{1}{2} x^T \Sigma_A^{-1} \cdot x$$

$$= \frac{1}{2}\left[x^T \cdot \Sigma_A^{-1} x - x^T \Sigma_B^{-1} \cdot x\right] = \underbrace{\frac{1}{2} x^T \cdot x}_{\frac{1}{2}\|x\|^2} \underbrace{\left(\Sigma_A^{-1} - \Sigma_B^{-1}\right)}$$

二次项 A

$P(A)$ $P(B)$

$P(C1)$ $P(C2)$ * data :

0.7 0.3

$$y = x^T \cdot A \cdot x + b^T \cdot x + c$$

$\Sigma_A^{-1} \neq \Sigma_B^{-1}$

$A \neq 0$

$A = 0$ $y = b^T \cdot x + c.$

$\Sigma_A = \sigma_A^2 I$

* 不一定 一定不相等

$\Sigma_B = \sigma_B^2 I$

$$\boxed{\Sigma_A = \Sigma_B = \sigma^2 I}$$

$$\underset{1\times p}{x^T} \underset{p\times p}{\Sigma^{-1}} \underset{p\times 1}{x} = \Sigma^{-1} x^T \cdot x$$

$$\neq x^T x \cdot \Sigma^{-1}$$

3.6个月　　125/天　100维度

125/2 = 62.5 $\begin{cases} \text{训} \\ \text{测} \end{cases}$

N个样本　P维

$N \geq P$

$N < P$ 过拟合

$N = 62.5$

$P = 100$

$N < P$

$\gamma =$

① $\begin{cases} W = (X^T \cdot X)^{-1} \cdot X^T \cdot t \\ W = (X^T \cdot X + \underset{②}{\gamma I})^{-1} \cdot X^T \cdot t \end{cases}$

正则化：

③ L1 : Lasso : $P(w) = \|w\|$: 权重稀疏

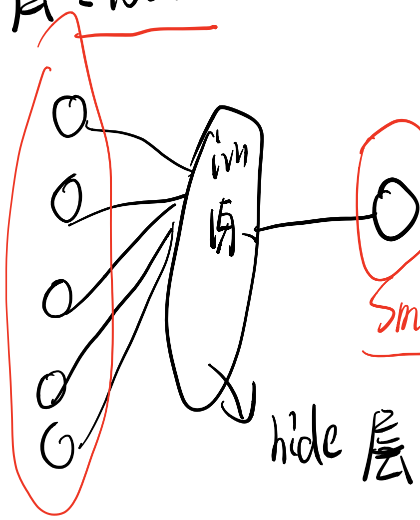
④ L2 : 岭回归 = 权重衰减 ⑤ $P(w) = w^T w = \|w\|^2$

$\arg\min \sum L(w) + \lambda P'(w)]$

4、 150 → 训练　　200 → data

50 → 测试　　15 → 维度

① hide 层 : 1000　(hidden_layer_sizes)



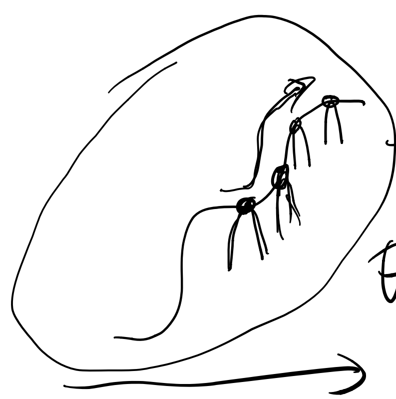
② max_iter = 2000000.

训练迭代

small values

4、 early_stopping = True ✓

5、EM



→ X全局

$\theta^{(t+1)} \geq \theta^{(t)}$

① MLE :

$\theta = \arg\max P(x|\theta)$

∂f = 0　$\hat{\theta}$ = max

$\frac{\partial}{\partial\theta} = 0$

② EM：

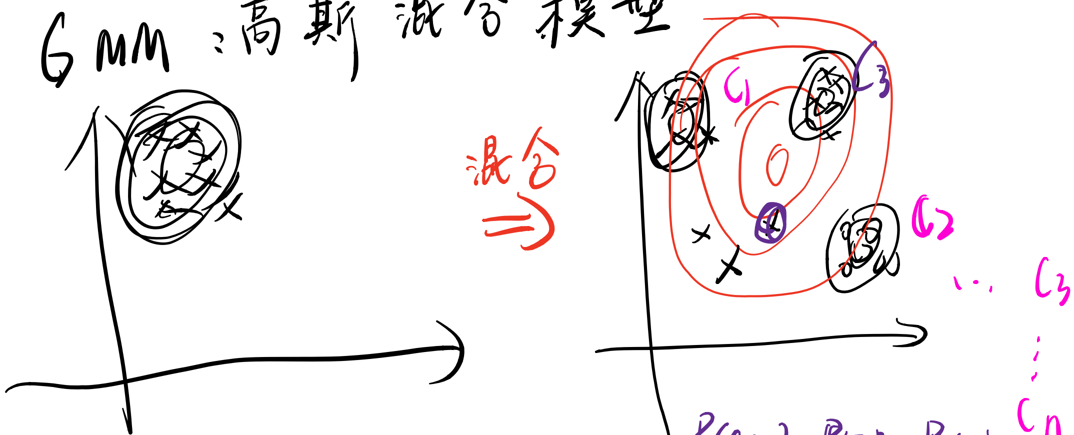
迭代：

$$\theta^{(t+1)} = \arg\max \int_z \log P(x, z \mid \theta)\, P(z \mid x, \theta^{(t)})\, dz$$

③ 收敛速度 → 取决于函数的性质

$\ln P_\theta(x)$

6、GMM：高斯混合模型



$P(C_1)$ $P(C_2)$ $P(C_3)$

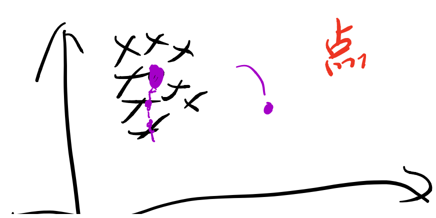
z	1	2	……	K
$p(z)$	P_1	P_2	……	P_k

软分类 $\begin{cases} P(C_1) \\ P(C_2) \end{cases}$ 标准

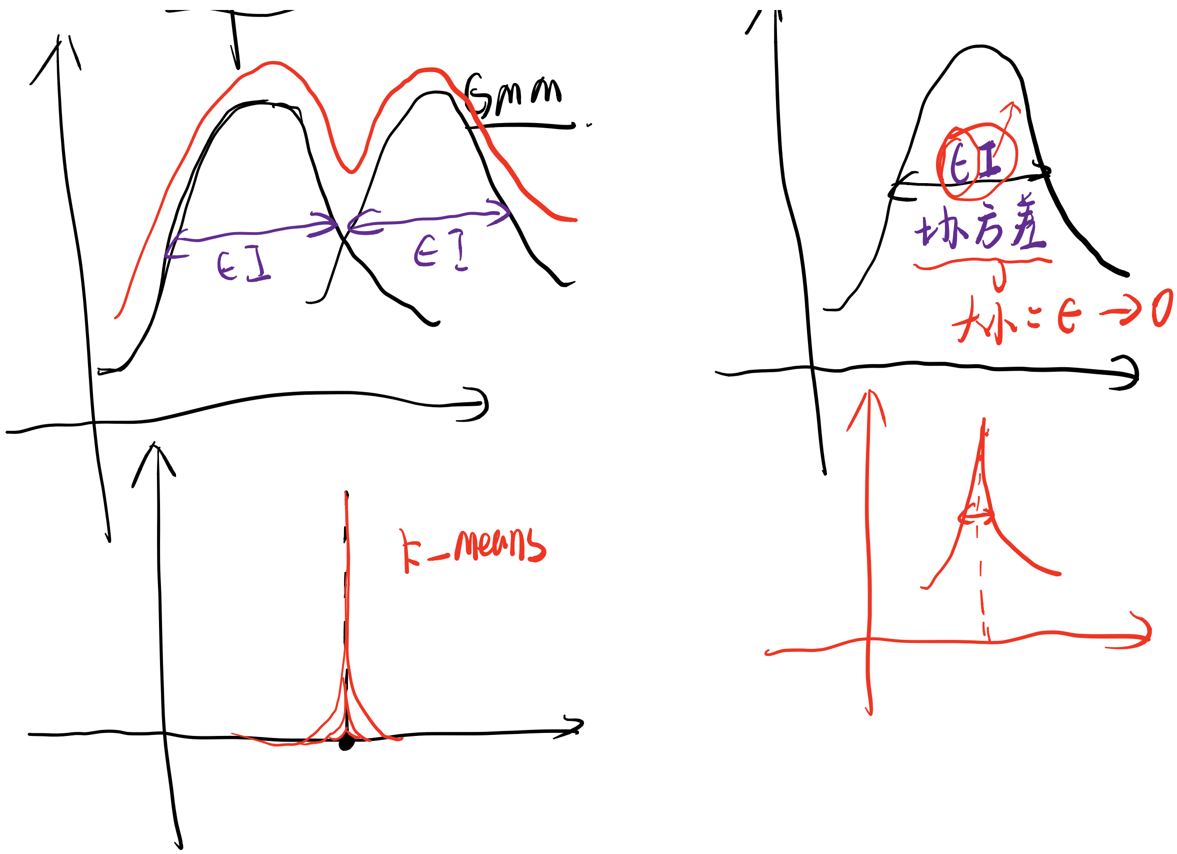
① $P(C_1) > P(C_2)$

② $\underset{0.7}{P(C_1)} - \underset{0.3}{P(C_2)} \geq 0.2$

K-means：硬分类 100%

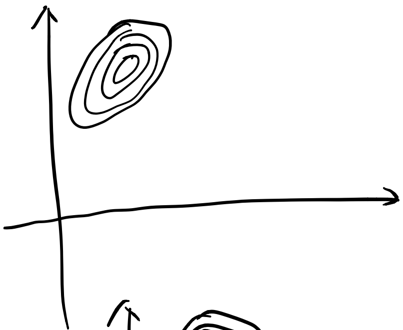


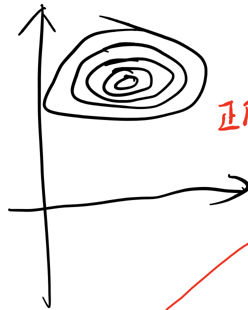
$\Sigma_1 = \dots \Sigma_k = \epsilon I$


GMM
ϵI
ϵI
k-means
ϵI
协方差
大小：ϵ→0

1、高斯分布:

① $X \sim N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{\Sigma}} \cdot \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right)$

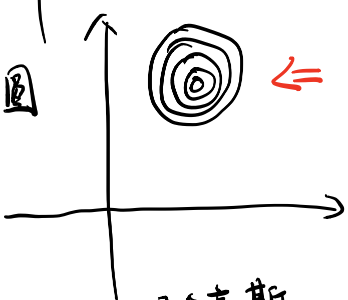




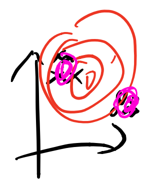
Σ: 对角矩阵 $\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix}$

正的椭圆.

圆



$\Leftarrow \lambda_1 = \lambda_2 = \lambda_p$



② GMM: 混合高斯

局限: 有的data 估计为一个高斯.

③ 边缘概率: $P(X_a)$ $P(X_b)$

条件概率: $P(X_a | X_b)$

联合概率: $P(X_a, X_b)$

③ $p(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)}$

④ $X \sim N(\mu, \Sigma)$

$y = AX + B$

$y \sim N(A\mu + B, A\Sigma A^T)$

用精度矩阵: Λ^{-1}

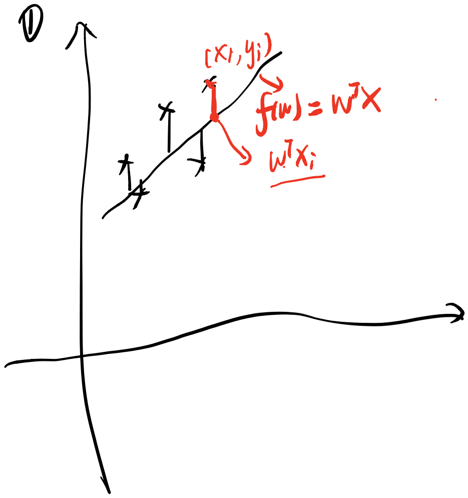
$y = AX + B + \varepsilon$ $X \sim N(\mu, \Lambda^{-1})$

$\varepsilon \sim N(0, L^{-1})$

$y \sim N(A\mu + B, L^{-1} + A\Lambda^{-1}A^T)$

2、线性回归（一次）

①



距离：（最小二乘估计）：

$$L(w) = \sum_{i=1}^{N} \| w^T x_i - y_i \|^2$$ Loss function.

$$= \sum_{i=1}^{N} (w^T x_i - y_i)^2$$ [矩阵的形式]

$$\hat{w} = \arg\min L(w)$$

$$\frac{\partial L(w)}{\partial w} = 2X^T \cdot X \cdot W - 2X^T \cdot Y = 0$$

最佳. $W = (X^T X)^{-1} X^T \cdot Y$

② Loss Function

$$L(w) = \sum_{i=1}^{N} \| w^T x_i - y_i \|^2$$

$X_{N \times P}$ N个样本 P维度. $N \gg P$

若不满足 $N \gg P$：过拟合

权值稀疏 L1：Lasso $P(w) = \|w\|$

L2：Ridge 岭回归 $P(w) = \|w\|^2 = w^T w$ 惩罚系数.

权值衰减

$$\arg\min_{w} [L(w) + \lambda P(w)] = J(w)$$

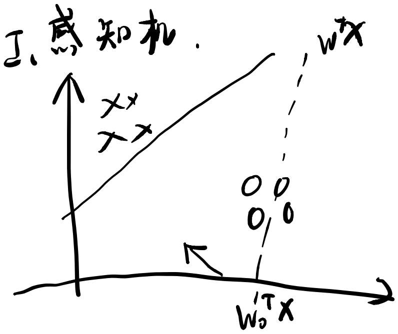
$$J(w) = w^T (X^T X + \lambda I) w - 2 w^T X^T Y + Y^T Y$$

$$\hat{w} = \arg\min_{w} J(w) = (X^T X + \lambda I)^{-1} X^T Y$$

3. 线性分类.

硬分类. $\begin{cases} \text{线性判别: fisher} \\ \text{感知机.} \end{cases}$

I. 感知机.



错误驱动.

$f(w) = \mathrm{sign}(w^Tx)$

$$\mathrm{sign}(a) = \begin{cases} +1 & a \geq 0 \\ -1 & a < 0 \end{cases}$$

$$\left.\begin{array}{ll} w^Tx_i > 0 & y_i = +1 \\ w^Tx_i < 0 & y_i = -1 \end{array}\right\} \rightarrow \text{分类正确.}$$

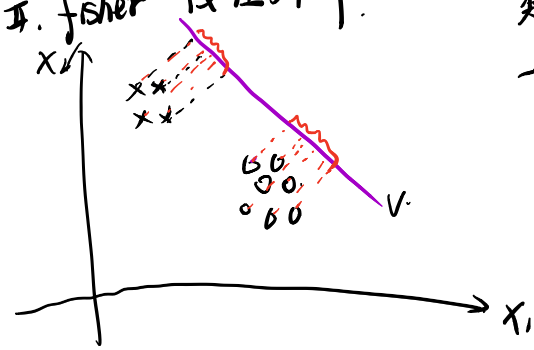
$\underline{y_i w^Tx_i > 0}$

Loss function

$$L(w) = \sum_{i=1}^{N} I\{y_i w^T \cdot x_i < 0\}.$$

$$L(w) = \sum_{\underline{x_i(-1)}} -y_i w^T x_i$$

II. fisher 线性判别.



类与类 均值 大

个类 方差 小

$$J(w) = \frac{(\bar{z}_1 - \bar{z}_2)^2 \uparrow\text{大}}{S_1 + S_2 \downarrow\text{小}}$$

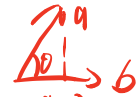
$$= \frac{[w^T(\bar{x}_{c_1} - \bar{x}_{c_2})]^2}{w^T(S_{c_1} + S_{c_2})w}$$

$$= \frac{w^T\underbrace{(\bar{x}_{c_1} - \bar{x}_{c_2})(\bar{x}_{c_1} - \bar{x}_{c_2})^T}_{S_B}w}{w^T\underbrace{(S_{c_1} + S_{c_2})}_{S_w}w}$$

$$= \frac{w^T \cdot S_B \cdot w}{w^T \cdot S_w w}$$

求最优解： $\frac{\partial J(w)}{\partial w} = 2S_b w(w^T S_w w)^{-1} + w^T S_b \cdot w(-1)(w^T \cdot S_w \cdot w)^{-2} \cdot 2S_w \cdot w$

$= 0$



$|\vec{a}||\vec{b}| \cdot \cos\theta$

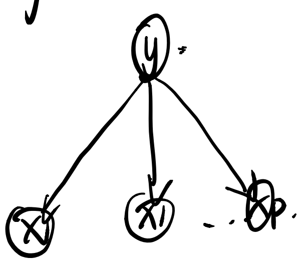
$a^T \cdot b$ ：投影

$$w = \frac{w^T \cdot S_w \cdot w}{w^T \cdot S_b \cdot w} \quad S_w^{-1} \cdot S_b \cdot w$$

$$(S_{c_1} + S_{c_2})^{-1} \cdot (\overline{x_{c_1}} - \overline{x_{c_2}}) \cdot \underbrace{(\overline{x_{c_1}} - \overline{x_{c_2}})^T \cdot w}_{实数}$$

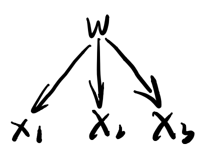
方向： $S_w^{-1} \cdot (\overline{x_{c_1}} - \overline{x_{c_2}})$

Ⅲ. Naive Bayes.



x_i x_j 相互独立.

$x_i \perp x_j \mid y. \quad (i \neq j)$

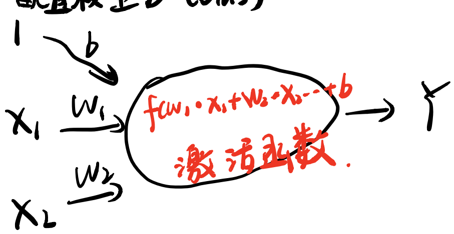


$$P(w|x_1) = \frac{P(x|w) \cdot P(w)}{P(x_1|w)P(w) + P(x_2|w)P(w) + P(x_3|w) \cdot P(w)}$$

4. 神经网络. MLP

Input layer

配置权重b (bias)



① Sigmoid (S型)：输入一个实值，输出一个0至1间值

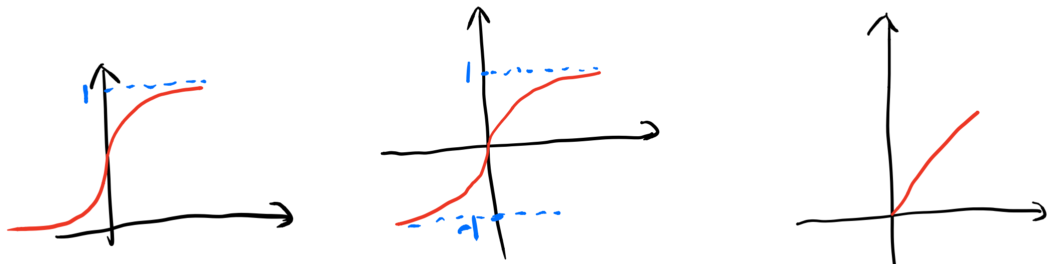
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

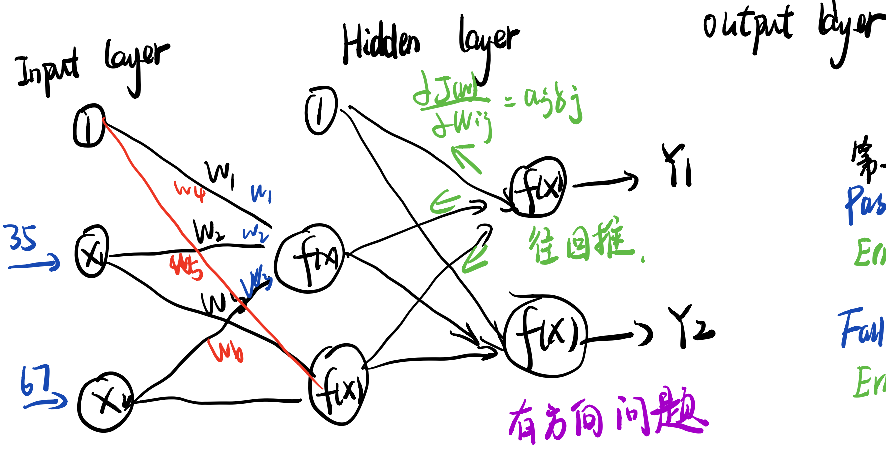
② tanh (双曲正切)：输入实值，输出一个[-1,1]间的值.

$$\tanh(x) = 2\sigma(2x) - 1$$

③ ReLU (修正线性单元)：输出实值，设定0的阈值

$$f(x) = \max(0, x)$$





Target=1

第一次
Pass = 0.4
Error = 1 - 0.4 = 0.6
Fail = 0.6
Error = 0 - 0.6 = 0.6

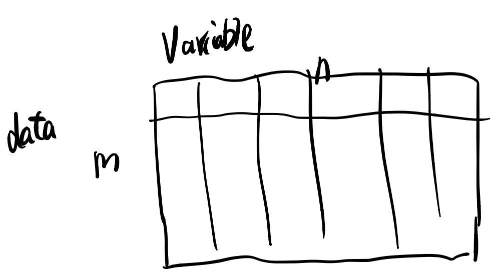
第二次
Pass = 0.8
Fail = 0.2

Target = 0

例：

Hour Study	Mid Term mark	Final Term Result
35	67	1
12	75	0
16	89	1
45	56	1
10	90	0

5. PCA 主成分提取



维度灾难：

$r=1 \quad V_{超球体} = K \cdot r^D$ D维度

$V_{外} = K \cdot r^D = K$

$V_{环} = K - K(1-\varepsilon)^D$

$$\frac{V_{环}}{V_{外}} = \frac{K - K(1-\varepsilon)^D}{K}$$

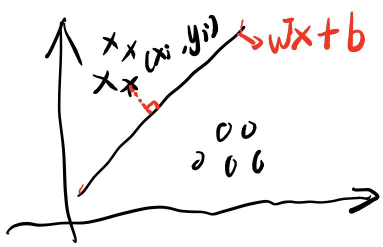
$$= 1 - (1-\varepsilon)^D \quad (0<\varepsilon<1)$$

$D \to \infty \quad \lim_{D\to\infty} (1-\varepsilon)^D = 0$

$$D \to \infty \quad \frac{V_{环}}{V_{外}} \Rightarrow 1$$

6. SVM 支持向量机

① $\begin{cases} \text{hard-margin SVM} \\ \text{soft-margin SVM} \\ \text{kernel SVM} \end{cases}$ 硬间隔 / 软 / 约束优化



目的：选出最好的一条

$f(w) = \mathrm{sign}(w^T x + b)$ 符号函数.

$\max \ \mathrm{margin}(w, b)$ 最大间隔分类器

$$\text{s.t.} \begin{cases} w^T x_i + b > 0 & y_i = +1 \\ w^T x_i + b < 0 & y_i = -1 \end{cases} \Rightarrow y_i(w^T x_i + b) > 0 \quad i = 1, 2, \dots N.$$

同号.

$$\text{distance} = \min_{\substack{w,b \\ x_i \\ i=1\dots N}} \frac{1}{\|w\|} |w^T x_i + b| = \mathrm{margin}(w, b)$$

$$\max \text{margin}(w,b) = \max \min \frac{1}{\|w\|} |w^T x_i + b|.$$

$$\text{s.t. } y_i (w^T x_i + b) > 0$$

$$\Rightarrow \begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w \ (\text{目标}\dots) \\ \text{s.t. } y_i (w^T x_i + b) \geq 0 \quad \text{for } \forall i = 1 \dots N \end{cases}$$ N个约束)

② 拉格朗日: 目标函数 + N条件

$$L(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \underbrace{\lambda_i}_{\geq 0} \underbrace{(1 - y_i (w^T x_i + b))}_{\leq 0}$$

令: $y_i (w^T x_i + b) = r \quad (r \geq 0)$

目的: 求极值.

③ 对偶问题:

max min distance $\leq$ min max distance

强对偶 $=$

弱对偶 $\leq$

7. EM

8. GMM

9. RoC : 算面积

AUC : 85% (具体分值) : 选出最好的分类器

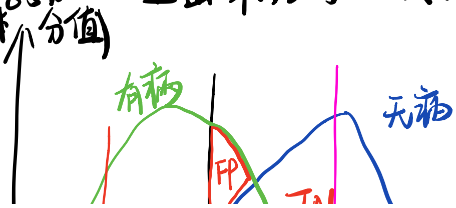
TN : True Nagtive　有病 无核

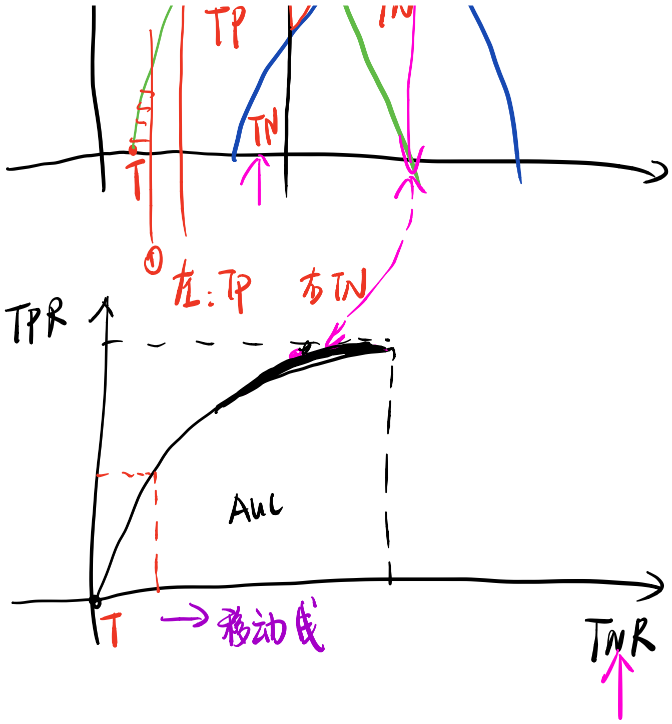
TP : True possitive : 检测=实际. P(病|核)

FN : False Nagtive　无病 无核

FP : False possitive　P(¬病|核)

检测≠实际.




TP
TN
TN
T
左:TP
右TN
TPR
AUC
T
移动线
TNR