SVM：支持向量机：hard margin，soft margin，kernel function
(线性可分SVM)　(线性SVM)　(解决线性不可分问题)(核函数)

① 线性可分向量机

<a> 过程概要

分割正负样本的超平面 $\vec{w}\vec{x}+b=0$

将样本 x_i 代入 $\vec{w}\cdot\vec{x}+b$，值
- >0：样本在超平面法线正向
- =0：在超平面上
- <0：在法线负方向

样本 x_i 到超平面距离 $\frac{|\vec{w}^T\vec{x}_i^T+b|}{\|\vec{w}\|}=d(x^{(i)},l_j)$ ☆

如果有样本的标注值(y值) {正样本：1，负样本：-1}

$$\frac{y^{(i)}\cdot(\vec{w}^Tx^{(i)T}+b)}{\|w\|}$$

预测正确：>0　正*正 or 负*负
预测错误：<0　正*负 or 负*正

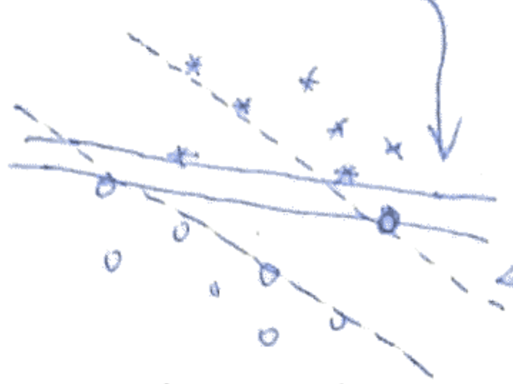
优化目标：找到超平面 l_j
使得 $\max_j(\min_{i\in[1,n]} d(x^{(i)},l_j))$
(样本到超平面的最小距离)→能够最大

<b> 建模

分离超平面：$w^T\Phi(x)+b=0$
其中 $\Phi(x)$ 是核函数，作特征空间转换，把x映射到更高维度

分类决策函数：$f(x)=\text{sign}(w^T\Phi(x)+b)$
- >0：在正的半空间
- <0：在负的半空间

<c> 线性可分向量机 v.s. 线性向量机



会引入松弛因子
放弃一部分样本
得到更大的分类间隔

<d> 求解线性可分SVM：用Lagrange乘子法

$$\arg\max_{w,b}\left\{\frac{1}{\|w\|}\min_i[\underbrace{y_i\cdot(w^T\cdot\Phi(x_i)+b)}_{\geq 1}]\right\}$$

不论值是多少，都可以通过等比缩放
$y_i(w^T\cdot\Phi(x_i)+b)*C$ ①
$\|w\|*C$ ②　来让 ①/② 值保持不变
$\boxed{\text{同时让 ①} \geq 1}$

C与求$\|w\|$最优解无关
可忽略

优化目标转为
让①取最小值1
(只关心支持向量)
忽略C

$$\arg\max_{w,b}\frac{1}{\|w\|}$$

问题转化为

求解 $\max_{w,b} \frac{1}{\|w\|}$ s.t. $y_i \cdot (w^T \cdot \Phi(x_i) + b) \geq 1$, $i=1,2,\cdots n$

等价于 $\min_{w,b} \frac{1}{2}\|w\|^2$ s.t. $y_i \cdot (w^T \cdot \Phi(x_i) + b) \geq 1$, $i=1,2,\cdots n$.

拉格朗日函数 $\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T \cdot \Phi(x_i) + b) - 1)$

拉格朗日乘子.

原问题： $\min_{w,b} \max_{\alpha} \mathcal{L}(w,b,\alpha)$ 求解最大值的最小值

需满足KKT条件

对偶问题： $\max_{\alpha} \min_{w,b} \mathcal{L}(w,b,\alpha)$ 求解最小值的最大值.

对 w, b 求偏导，令偏导为 0

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2} \cdot 2w - \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) \to 0 \Rightarrow \boxed{w = \sum_{i=1}^{n} \alpha_i \cdot y_i \Phi(x_i)}$$ ~~求解出~~

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i \to 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

代入拉格朗日函数

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\boxed{\|w\|^2} - \sum_{i=1}^{n} \alpha_i (y_i (w^T \cdot \Phi(x_i) + b) - 1)$$

$$= \frac{1}{2} w^T \underbrace{\boxed{w}}_{代入} - w^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) - b \underbrace{\boxed{\sum_{i=1}^{n} \alpha_i y_i}}_{=0} + \sum_{i=1}^{n} \alpha_i$$

$$= \frac{1}{2} \underbrace{\boxed{w^T}}_{代入} \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) - w^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) + \sum_{i=1}^{n} \alpha_i$$

$$= -\frac{1}{2} \left(\sum_{i=1}^{n} \alpha_i y_i \Phi(x_i)\right)^T \left(\sum_{i=1}^{n} \alpha_i y_i \Phi(x_i)\right) + \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \Phi^T(x_i) \Phi(x_j)$$ //已经只有样本和 α_i 了 可以求解 α_i 了

根据对偶问题求max

$$\alpha_i^* = \arg\max_{\alpha_i} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \Phi^T(x_i) \Phi(x_j)\right)$$

求解 $\max_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\Phi(x_i) \cdot \Phi(x_j))\right)$ s.t. $\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \alpha_i \geq 0$

等价于 $\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\Phi(x_i) \cdot \Phi(x_j)) - \sum_{i=1}^{n} \alpha_i\right)$ s.t. $\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \alpha_i \geq 0$

对 α_i 求偏导，梯度下降.

也可以用SMO算法，固定任意 n-2个 α_i，只变化其中2个，让其下降更快，不断迭代，序列化求解

解出 α_i^* 之后，代入公式解出 w^*，b^*（只有支持向量经过的样本 $\alpha_i \neq 0$）

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i \Phi(x_i)$$ 上一页公式⊗

$$b^* = y_i - \sum_{i=1}^{N} \alpha_i^* y_i (\Phi(x_i)\cdot\Phi(x_j))$$ ③ ⟶ 代入支持向量算出？

代入公式得到

分离超平面 $w^*\Phi(x) + b^* = 0$

分类决策函数 $f(x) = \mathrm{sign}(w^*\Phi(x) + b^*)$

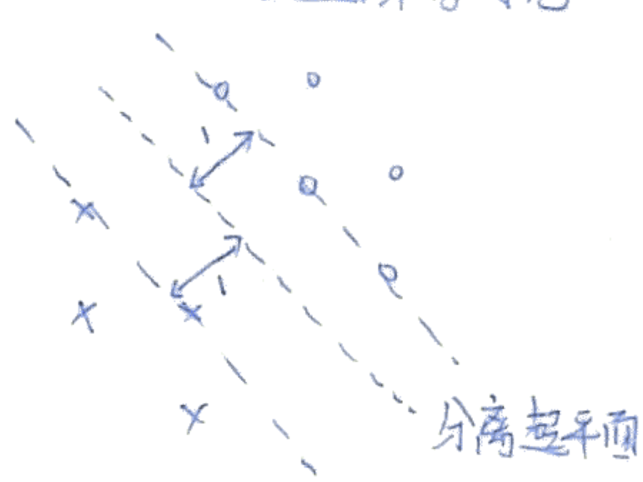
<e> 例子：见PPT

上一小节的算法

② 线性 SVM：$\boxed{\text{线性可分 SVM（① <d>）}}$中，希望 $\min\limits_{i\in[1,N]} d(x^{(i)}, \underbrace{L_j}_{\text{超平面}}) = \min\limits_{i\in[1,N]} \frac{|W^T\Phi(x_i)^T+b|}{\|W\|}$ 取最大

考虑样本标注时，就是期望 $\min\limits_{i\in[1,N]} \frac{y^{(i)}\cdot(W^T\Phi(x^{(i)})^T+b)}{\|W\|}$ 取最大

不妨标记 $\min\limits_{i\in[1,N]} \frac{y^{(i)}(W^T\Phi(x^{(i)})^T+b)}{\|W\|} = d_0$



$$\underbrace{y^{(i)}}_{\text{取值}+1,-1}(W^T\Phi(x^{(i)})^T+b) \geq d_0\|W\|$$

对 W 等比缩放，总能找到一个 W'，使得

$$y^{(i)}(W'^T\Phi(x^{(i)})^T+b) \geq 1$$

而在 $\boxed{\text{线性SVM}}$（本节的算法），为每个样本 x_i 指定一个松弛因子 ξ_i，约束条件变为

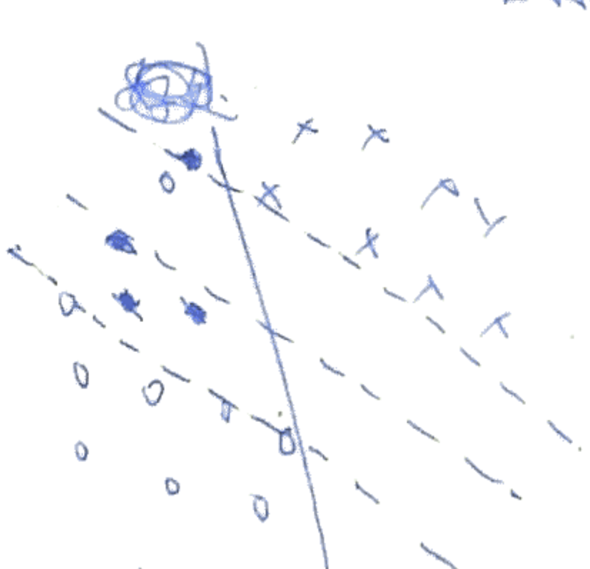
$$\boxed{y^{(i)}(W'^T\Phi(x^{(i)})^T+b) \geq 1-\xi_i}$$

目标函数也由 $\min\limits_{w,b}\frac{1}{2}\|W\|^2$ 变为

$$\boxed{\min_{w,b}\left(\frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\xi_i\right)}$$

权重

松弛因子也作损失值，加入到优化目标中，期望其最小



好处是有更宽的分类间隔
受噪声的影响小

作用类似于LR，线性回归的正则项

(a) 加入松弛因子以后的约束条件下求极值问题

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \quad s.t. \begin{cases} y_i(w\cdot x_i + b) \geq 1-\xi_i & i=1,2,\cdots n \\ \xi_i \geq 0 & i=1,2,\cdots n \end{cases}$$

值越小，分类间隔宽度越大，抗过拟合越强

拉格朗日函数

$$\mathcal{L}(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(wx_i+b)-1-\xi_i) - \sum_{i=1}^{n}\mu_i\xi_i$$

对 w, b, ξ 求偏导，令偏导为0，得到

$$w = \sum_{i=1}^{n}\alpha_i y_i \Phi(x_n)$$

$$0 = \sum_{i=1}^{n}\alpha_i y_i$$

$$C - \alpha_i - \mu_i = 0$$

原问题：$\min\limits_{w,b,\xi}\max\limits_{\alpha} \mathcal{L}(w,b,\xi,\alpha,\mu)$

对偶问题：$\max\limits_{\alpha}\min\limits_{w,b,\xi} \mathcal{L}(w,b,\xi,\alpha,\mu)$

参考①的推导

将三式代入 $\mathcal{L}(w,b,\xi,\alpha,\mu)$，得到

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j) + \sum_{i=1}^{n}\alpha_i$$

接下来对 α 求 max

$$\max_{\alpha}\min_{w,b,\xi} \mathcal{L}(w,b,\xi,\alpha,\mu) = \max_{\alpha}\left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j) + \sum_{i=1}^{n}\alpha_i\right)$$

$$s.t. \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\begin{cases} C - \alpha_i - \mu_i = 0 \\ \alpha_i \geq 0 \\ \mu_i \geq 0 \quad i=1,2,\cdots n \end{cases} \Longleftrightarrow \boxed{0 \leq \alpha_i \leq C}$$

已从公式中消去

把负号提出，整理约束条件，上式等价于求

$$\min_{\alpha}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j) - \sum_{i=1}^{n}\alpha_i\right) \quad s.t. \begin{cases} C-\alpha_i-\mu_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$

对 α_i 求偏导，梯度下降，得到最优解 α^*（与线性可分SVM不同，α取各种值）

回代到公式②中，得到 $w^* = \sum_{i=1}^{n}\alpha_i^* y_i x_i$

使用满足 $0<\alpha_j<C$ 的向量，实践中往往取支持向量的均值作为 b^*

$$\rightarrow b^* = \frac{1}{2}\left(\max_{i:y=-1} w^* x_i + \min_{i:y=1} w^*\cdot x_i\right)$$

负例　　正例

求得分离超平面 $w^*x + b^* = 0$

求得分类决策函数 $f(x) = sign(w^*x + b^*)$

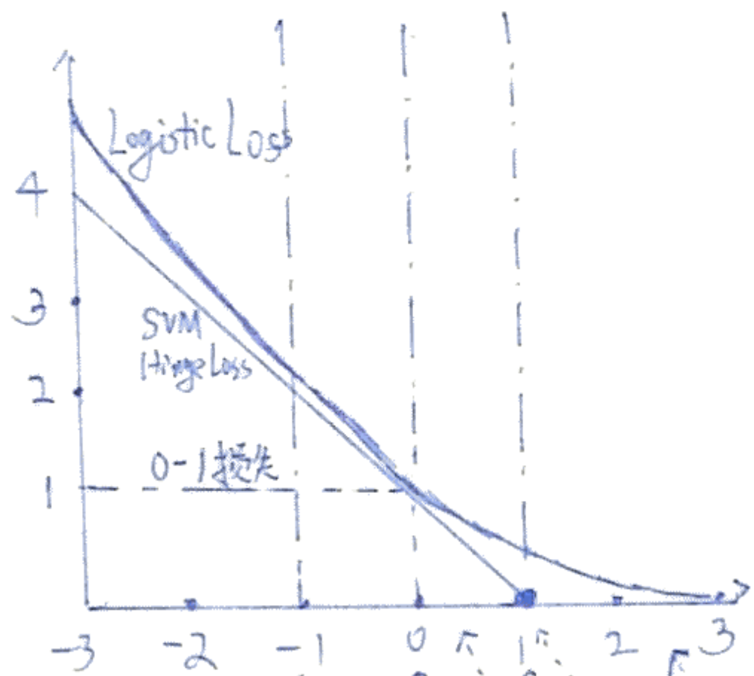
<5> 线性SVM损失函数分析

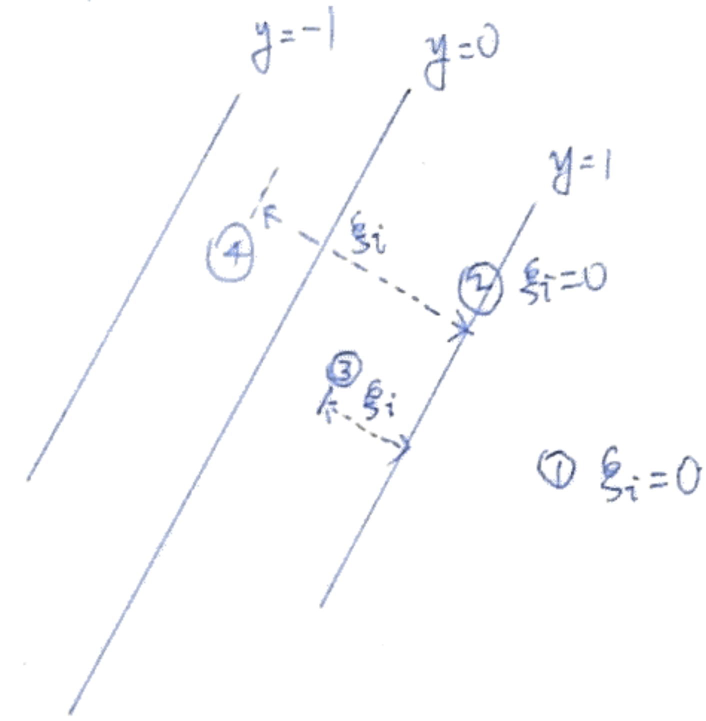
x_i约束条件的系数，即 $y_i(W\cdot x_i+b)\geq 1-\xi_i$ 的拉格朗日乘子

ξ_i系数，即 $\xi_i\geq 0$ 的拉格朗日乘子

前面推导过程中用到过约束条件 $C=\alpha_i+\mu_i$, $\mu_i\geq 0$

得到 $\begin{cases}\mu_i\geq 0\\ \alpha_i=C-\mu_i\in[0,C]\end{cases}$





样本① 分对

x_i在支撑面正上方：$\alpha_i=0$，$\xi_i=0$，(不需松弛)

(不是支撑向量，SVM不关心)

样本② x_i在分对支撑面上：$0<\alpha_i<C$，$\xi_i=0$

(是支撑向量) (不需松弛)

样本③ x_i在分隔超平面，分对支撑面之间

$\alpha_i=C$；$0<\xi_i<1$

(需松弛)

因为 $\mu_i\xi_i=0$ 而 $\xi_i>0$

$\Rightarrow \mu_i$必须为0

$\Rightarrow \alpha_i=C-\mu_i=C$

样本④ x_i在分隔超平面负方向

$\alpha_i=C$; $\xi_i>1$

推导同样本③ 如图.

0-1损失：分对无损失，分错损失值为1（每个样本）

Logistic Loss：分对也会有可能有一定损失，只是损失值小 (即使在支撑面正方向)

SVM Hinge Loss：支撑面正方向就无损失

分对但落下支撑面下方(需松弛)有数<1的损失

分错有>1的损失，错得越离谱损失越大

(但是线性的)

③ 核函数：将原始特征映射到新的特征空间，解决线性不可分问题

e.g. $\Phi(x_1, x_2, x_3) = \begin{pmatrix} x_1x_1 \\ x_1x_2 \\ \vdots \\ x_2x_1 \\ \vdots \\ x_3x_2 \\ x_3x_3 \\ \sqrt{2c}\,x_1 \\ \sqrt{2c}\,x_2 \\ \sqrt{2c}\,x_3 \\ c \end{pmatrix}$

多项式核函数（以 d=2 为例）：$\Phi(\vec{x}) = \left(\underbrace{vec}_{\text{数组}}(x_i x_j)\big|_{i,j=1}^{n},\ vec(\sqrt{2c}\,x_i)\big|_{i=1}^{n},\ c \right)$

推导来源 $k(\vec{x}, \vec{y}) = (\vec{x}\cdot\vec{y} + c)^2 = (\vec{x}\cdot\vec{y})^2 + 2c\vec{x}\vec{y} + c^2$

高维空间映射

$= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i x_j)(y_i y_j) + \sum_{i=1}^{n}(\sqrt{2c}\,x_i \cdot \sqrt{2c}\,x_j) + c^2$

d 值越大，精度越高，抗过拟合越弱

$= \Phi(\vec{x})^T \Phi(\vec{y})$ 导出

高斯核函数：$\Phi(\vec{x}) = e^{-\frac{x^2}{2\sigma^2}}\left[1, \sqrt{\frac{1}{1!}}\frac{x}{\sigma}, \sqrt{\frac{1}{2!}}\frac{x^2}{\sigma^2}, \sqrt{\frac{1}{3!}}\frac{x^3}{\sigma^3}, \cdots, \sqrt{\frac{1}{n!}}\frac{x^n}{\sigma^n}, \cdots\right]$

映射到了无穷维
理论上不可分
实际用高斯核
RBF 可分

推导来源 $k(\vec{x}, \vec{y}) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} = e^{-\frac{(x-y)^2}{2\sigma^2}} = e^{-\frac{x^2+y^2}{2\sigma^2}} \cdot \boxed{e^{\frac{xy}{\sigma^2}}}$ 泰勒展开

$= e^{-\frac{x^2+y^2}{2\sigma^2}}\left(1 + \frac{1}{\sigma^2}\frac{xy}{1!} + \left(\frac{1}{\sigma^2}\right)^2\frac{(xy)^2}{2!} + \left(\frac{1}{\sigma^2}\right)^3\frac{(xy)^3}{3!} + \cdots + \left(\frac{1}{\sigma^2}\right)^n\frac{(xy)^n}{n!} + \cdots\right)$

$= e^{-\frac{x^2+y^2}{2\sigma^2}}\left(1\cdot 1 + \frac{1}{1!}\frac{x}{\sigma}\frac{y}{\sigma} + \frac{1}{2!}\frac{x^2 y^2}{\sigma^2\sigma^2} + \frac{1}{3!}\frac{x^3}{\sigma^3}\frac{y^3}{\sigma^3} + \cdots + \frac{1}{n!}\frac{x^n}{\sigma^n}\frac{y^n}{\sigma^n} + \cdots\right)$

$= \Phi(x)^T \Phi(Y)$ 导出

Sigmoid 核函数：推导来源 $K(x, y) = \tanh(x \cdot y + c)$

选择核函数：往往要依赖先验领域知识 / 交叉验证等方案

没有先验知识时可以无脑使用高斯核，发现过拟合再退回到多项式核

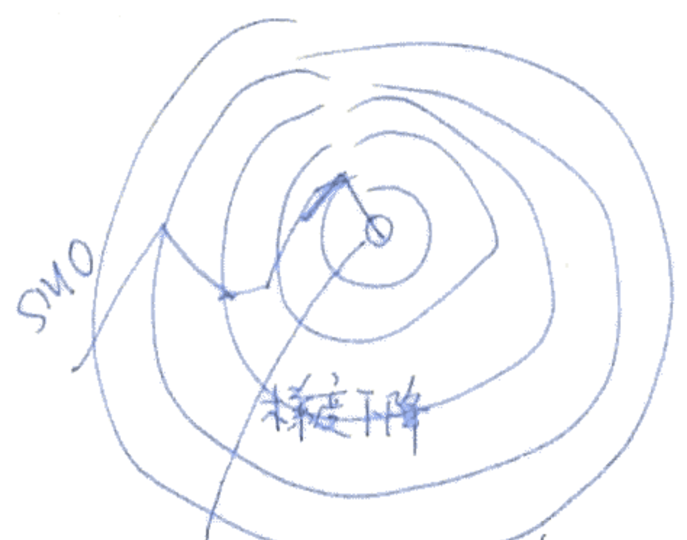
④ SVM SMO (Sequential Minimal Optimization)：序列最小最优化参数求解

SVM 参数求解时，有多个拉格朗日乘子，每次只选择其中两个乘子做优化，

其他因子认为是常数（将 N 个解问题，转成两个变量求解问题）

求解这个函数时

$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i$



s.t. $\sum_{i=1}^{N}\alpha_i y_i = 0$

$0 \le \alpha_i \le C$

改为求解

$\min_{\alpha_1\alpha_2} W(\alpha_1, \alpha_2) = \frac{1}{2}K_{11}\alpha_1^2 + \frac{1}{2}K_{22}\alpha_2^2 + y_1 y_2\alpha_1\alpha_2 K_{12} - (\alpha_1 + \alpha_2) + y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_i K_{i1} + y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_i K_{i2}$

s.t. $\alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{N} y_i\alpha_i = \varsigma \quad 0 \le \alpha_i \le C$

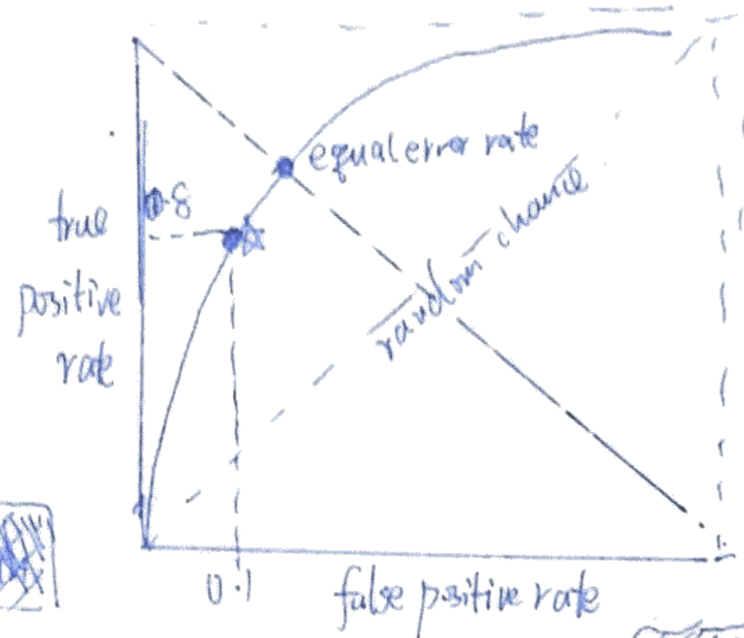
⑤ SVR：SVM用于回归

⑥ SVM用于多分类（How?）

⑦ AUC

预测 \ 实际	Positive	Negtive
正	TP ✓	FN ×
负	FP ×	TN ✓

recall



$$TPR = \frac{TP}{TP+FN}$$
(true positive rate) 实际是正的

$$FPR = \frac{FP}{FP+TN}$$
(false positive rate) 实际是负的

☆（FPR = 0.1，TPR = 0.8）：
以0.1的负例预测为正例的代价（10%）
换来0.8的正例预测为正例的效果（80%）

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP}$$

$$precision = \frac{TP}{TP+FP}$$

预测为正的样本中有百分之多少是正确预测的

$$recall = \frac{TP}{TP+FN}$$

总的正样本中，有多少能被预测出来

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

$$F_\beta = \frac{(1+\beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}$$