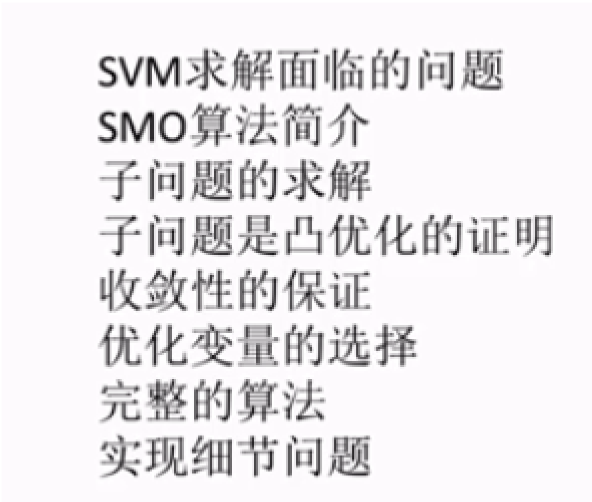
15 支持向量机2

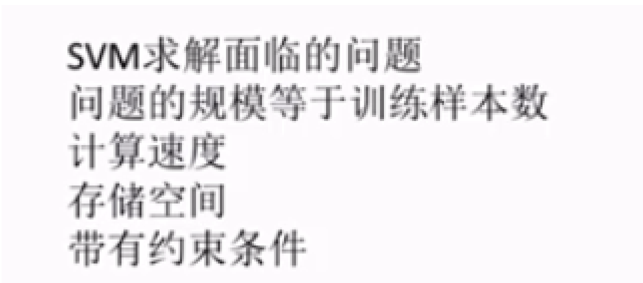
15.1 本集内容简介

上节课三种情况，线性可分，线性不可分，

为什么使用SMO算法



15.2 对偶问题求解面临的问题



正定二次型？

回忆SVM的对偶问题

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(\mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j\right)-\sum_{k=1}^{l}\alpha_k$$

$$0 \le \alpha_i \le C$$

$$\sum_{j=1}^{l}\alpha_j y_j = 0$$

$$\min_{\alpha} \frac{1}{2}\alpha^{\mathrm{T}}\mathrm{Q}\alpha - \mathrm{e}^{\mathrm{T}}\alpha$$

$$\mathrm{y}^{\mathrm{T}}\alpha = 0$$

$$0 \le \alpha_i \le C, i = 1,\dots,l$$

$l \times l$

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(\mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j\right)-\sum_{k=1}^{l}\alpha_k$$

$$0 \le \alpha_i \le C$$

$$\sum_{j=1}^{l}\alpha_j y_j = 0$$

$(\alpha_1 \alpha_2 \cdots \alpha_l)$

$$\min_{\alpha} \frac{1}{2}\alpha^{\mathrm{T}}\mathrm{Q}\alpha - \mathrm{e}^{\mathrm{T}}\alpha$$

$e: (1, \dots, 1)$

$$\mathrm{y}^{\mathrm{T}}\alpha = 0$$

$$0 \le \alpha_i \le C, i = 1,\dots,l$$

二次规划

15.3 SMO算法简介

SMO算法简介

Sequential minimal optimization，顺序最小优化

Platt等人于1998年提出

核心思想是分治法，每次挑选出两个变量进行优化

$$K(\mathrm{x}_i,\mathrm{x}_j)=\phi(\mathrm{x}_i)^{\mathrm{T}}\phi(\mathrm{x}_j) \qquad K_{ij}=K(\mathrm{x}_i,\mathrm{x}_j)$$

$$Q_{ij}=y_i y_j K(\mathrm{x}_i,\mathrm{x}_j) \qquad Q_{ij}=y_i y_j K_{ij}$$

$$u_i=\sum_{j=1}^{l} y_j \alpha_j K(\mathrm{x}_j,\mathrm{x}_i)-b$$

最优点处必须满足

$$\alpha_i=0 \Leftrightarrow y_i u_i \ge 1$$

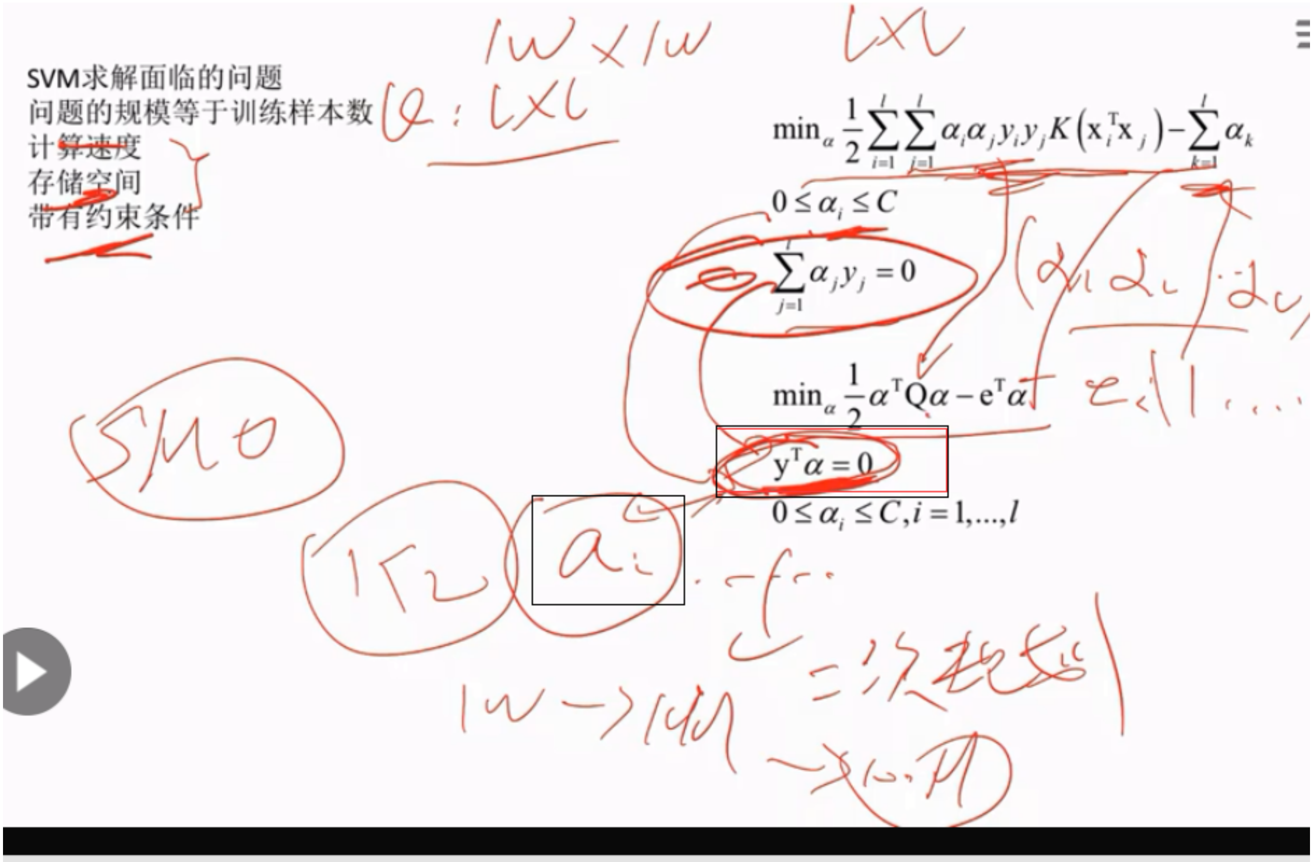
$$0<\alpha_i<C \Leftrightarrow y_i u_i=1$$

$$\alpha_i=C \Leftrightarrow y_i u_i \le 1$$

下图中

图中举行方框的等式约束，如果alpha i 调整不等于0。

最少要挑选两个变量来优化。



再来看变量的定义

- 定义核矩阵
- 定义Q
- 定义ui

$$K(\mathrm{x}_i,\mathrm{x}_j)=\phi(\mathrm{x}_i)^{\mathrm{T}}\phi(\mathrm{x}_j) \qquad \boxed{K_{ij}=K(\mathrm{x}_i,\mathrm{x}_j)}$$

$$Q_{ij}=y_i y_j K(\mathrm{x}_i,\mathrm{x}_j) \qquad \boxed{Q_{ij}=y_i y_j K_{ij}}$$

$$\boxed{u_i=\sum_{j=1}^{l}y_j\alpha_j K(\mathrm{x}_j,\mathrm{x}_i)-b}$$

最优点处必须满足

$$\alpha_i=0 \Leftrightarrow y_i u_i \ge 1$$
$$\boxed{0<\alpha_i<C \Leftrightarrow y_i u_i = 1}$$
$$\alpha_i=C \Leftrightarrow y_i u_i \le 1$$

15.4 求解子问题

两个变量子问题

$$f\left(\alpha_i,\alpha_j\right)=\frac{1}{2}K_{ii}\alpha_i^2+\frac{1}{2}K_{jj}\alpha_j^2+sK_{ij}\alpha_i\alpha_j+y_iv_i\alpha_i+y_jv_j\alpha_j-\alpha_i-\alpha_j+c$$

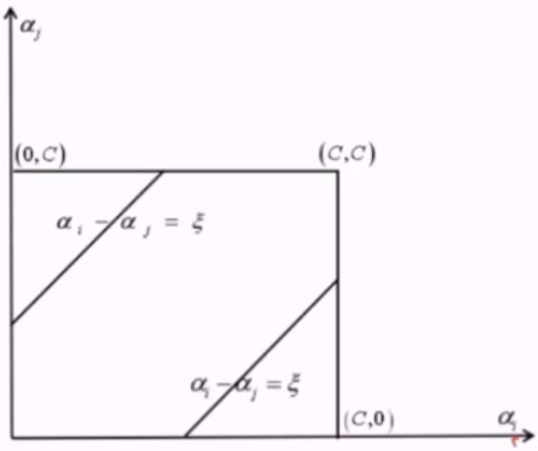
$$s=y_iy_j$$

$$v_i=\sum_{k=1,k\neq i,k\neq j}^{l}y_ka_k^*K_{ik}$$

$$0\le\alpha_i\le C$$
$$0\le\alpha_j\le C$$

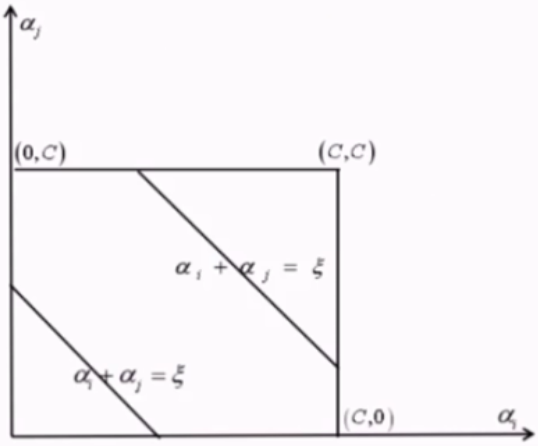
$$y_i\alpha_i+y_j\alpha_j=-\sum_{k=1,k\neq i,k\neq j}^{l}y_k\alpha_k=\xi$$

$\alpha_i-\alpha_j=\xi$



$$L=\max\left(0,\alpha_j-\alpha_i\right)$$
$$H=\min\left(C,C+\alpha_j-\alpha_i\right)$$

$\alpha_i+\alpha_j=\xi$



$$L=\max\left(0,\alpha_j+\alpha_i-C\right)$$
$$H=\min\left(C,\alpha_j+\alpha_i\right)$$

前面讲到的，

两个变量子问题

$(\alpha_i\ \alpha_j) \to l$

$$f(\alpha_i,\alpha_j)=\frac{1}{2}K_{ii}\alpha_i^2+\frac{1}{2}K_{jj}\alpha_j^2+sK_{ij}\alpha_i\alpha_j+y_iv_i\alpha_i+y_jv_j\alpha_j-\alpha_i-\alpha_j+c$$

$$s=y_iy_j$$

$$v_i=\sum_{k=1,k\neq i,k\neq j}^{l} y_k a_k^* K_{ik}$$

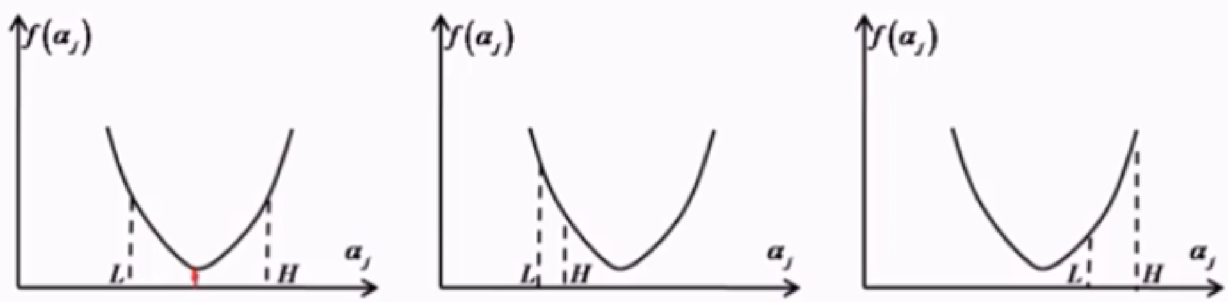
$$0\le\alpha_i\le C$$

$$0\le\alpha_j\le C$$

$$y_i\alpha_i+y_j\alpha_j=-\sum_{k=1,k\neq i,k\neq j}^{l} y_k\alpha_k=\xi$$

$+1\ +1$

$$\alpha_j^{new}=\begin{cases}H & \alpha_j^{new,unclipped}>H\\ \alpha_j^{new,unclipped} & L\le\alpha_j^{new,unclipped}\le H\\ L & \alpha_j^{new,unclipped}<L\end{cases}$$



$f(a_j) \to$

$[L, H]$

$$\alpha_i + y_i y_j \alpha_j = y_i \xi \qquad \alpha_i = y_i \xi - y_i y_j \alpha_j$$

$$w = y_i \xi$$

$$\frac{1}{2} K_{ii} \left(w - s\alpha_j\right)^2 + \frac{1}{2} K_{jj} \alpha_j^2 + sK_{ij} \left(w - s\alpha_j\right)\alpha_j + y_i v_i \left(w - s\alpha_j\right) + y_j v_j \alpha_j - \left(w - s\alpha_j\right) - \alpha_j + c$$

求导并令导数为0

$$K_{ii}\left(w - s\alpha_j\right)(-s) + K_{jj}\alpha_j + sK_{ij}\left(w - 2s\alpha_j\right) - s y_i v_i + y_j v_j + s - 1 = 0$$

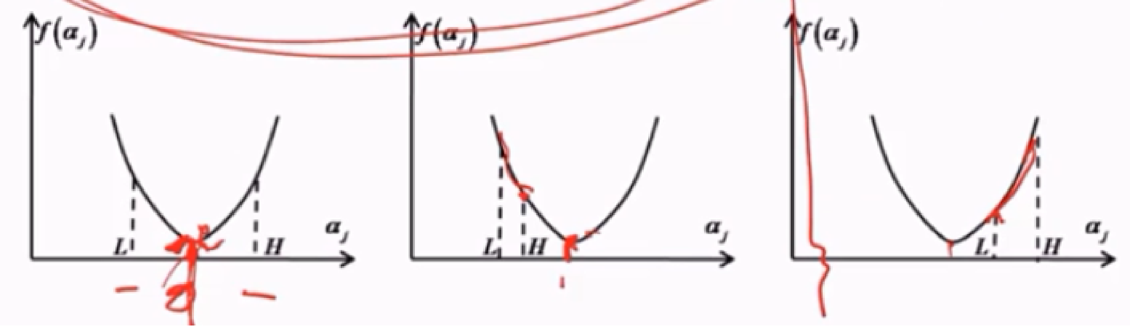
$$s y_i v_i = y_i y_j y_i v_i = y_j v_i$$

$$\left(K_{ii} + K_{jj} - 2K_{ij}\right)\alpha_j - swK_{ii} - swK_{ij} - y_j v_i + y_j v_j + s - 1 = 0$$

$$\left(K_{ii} + K_{jj} - 2K_{ij}\right)\alpha_j = sw\left(K_{ii} + K_{ij}\right) + y_j v_i - y_j v_j + 1 - s$$

$$\left(K_{ii} + K_{jj} - 2K_{ij}\right)\alpha_j = \alpha_j^*\left(K_{ii} + K_{jj} - 2K_{ij}\right) + y_j\left(u_i - u_j + y_j - y_i\right)$$

$$\alpha_j^{new} = \begin{cases} H & \alpha_j^{new,unclipped} > H \\ \alpha_j^{new,unclipped} & L \le \alpha_j^{new,unclipped} \le H \\ L & \alpha_j^{new,unclipped} < L \end{cases}$$



$$\alpha_j^{new,clipped} = \begin{cases} H & \alpha_j^{new} > H \\ \alpha_j^{new} & L \le \alpha_j^{new} \le H \\ L & \alpha_j^{new} < L \end{cases}$$

$$\alpha_i^{new} = \alpha_i + s\left(\alpha_j - \alpha_j^{new,clipped}\right)$$

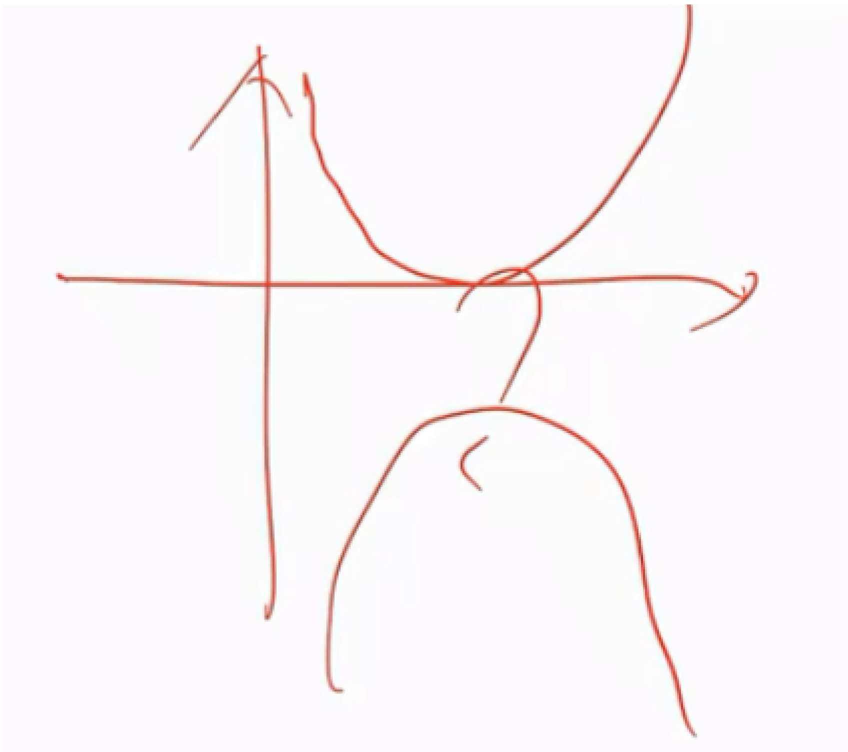
15.5 子问题是凸优化问题的证明

子问题是一个凸优化问题

$$\begin{bmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{bmatrix}$$

$$\begin{bmatrix} y_i \mathrm{x}_i^{\mathrm{T}} \\ y_i \mathrm{x}_i^{\mathrm{T}} \end{bmatrix} \left[y_i \mathrm{x}_i, y_i \mathrm{x}_j \right] = \mathrm{A}^{\mathrm{T}} \mathrm{A}$$

前面说，抛物线是开口向上的。



是一个凸优化问题，对于一个二元二次矩阵，它的汉森矩阵是正定的。

根据Q的定义，

$$Q_{ij} = y_i y_j (x_i^T x_j)$$

写成矩阵的形式

$$\begin{bmatrix} y_i \mathbf{x}_i^{\mathrm{T}} \\ y_i \mathbf{x}_i^{\mathrm{T}} \end{bmatrix} \left[y_i \mathbf{x}_i, y_i \mathbf{x}_j \right] = \mathbf{A}^{\mathrm{T}} \mathbf{A}$$

等价于核函数的映射。

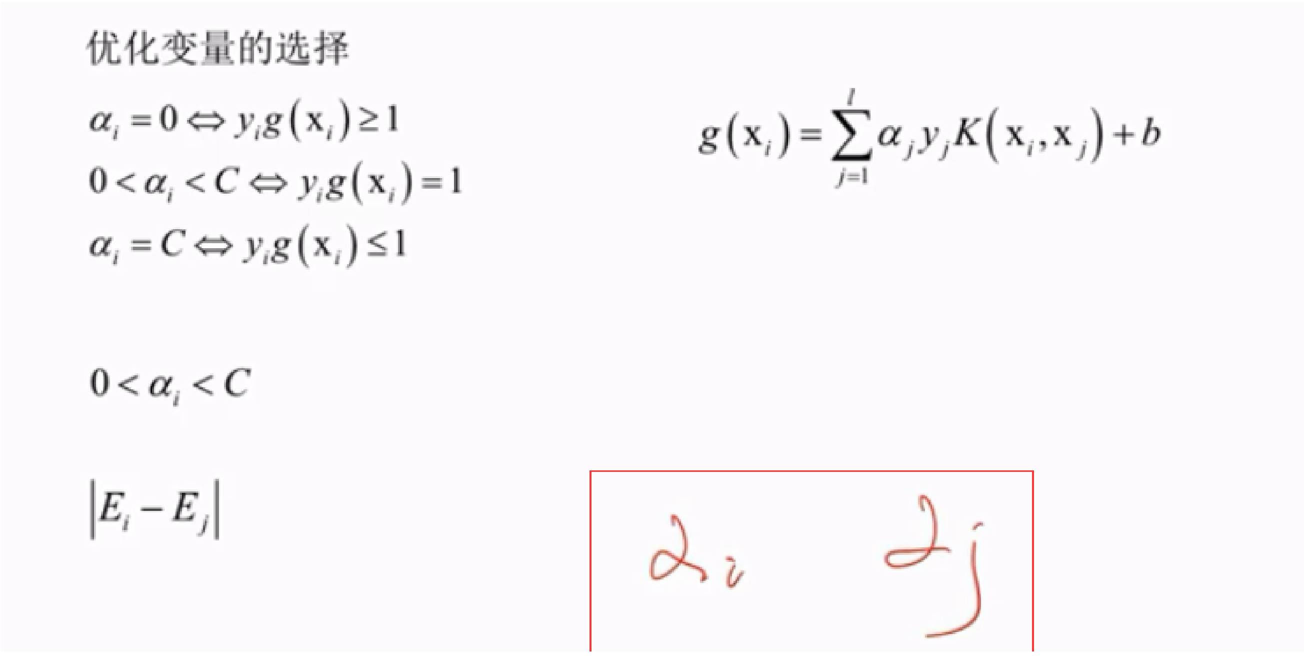
支持向量机，是一个凸优化问题。

15.6 SMO算法的收敛性

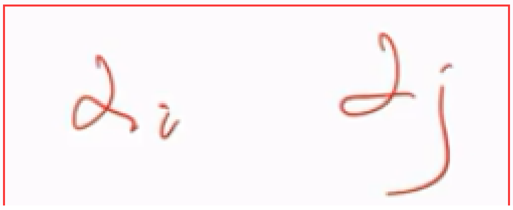
收敛性的保证

无论本次迭代时两个变量的初始值是多少，通过上面的子问题求解算法得到是在可行域里的最小值，因此每次求解更新这两个变量的值之后，都能保证目标函数值小于或者等于初始值，即函数值下降，所以SMO算法能保证收敛

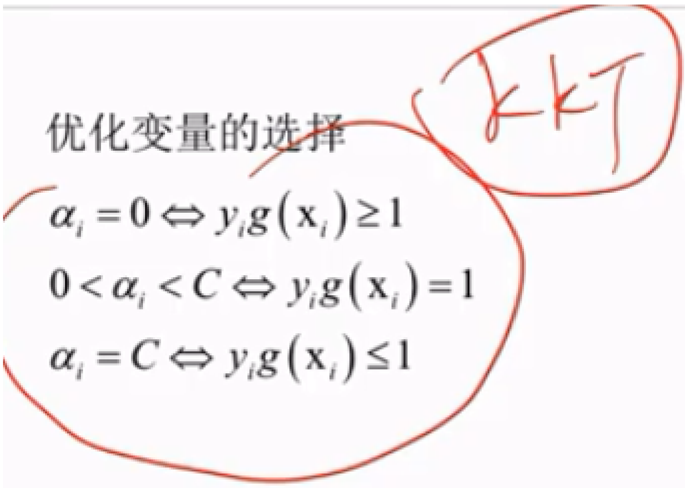
15.7 优化变量的选择



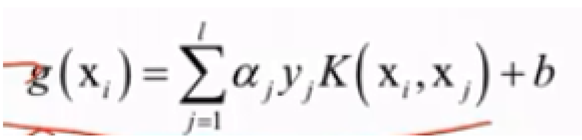
核心问题是找出alpha i 和alpha j



在最优解的时候，必须满足KTT条件



反过来说，如果还没有满足KTT条件，那么就是还没有达到最优解。



把xi 带进

优化变量的选择

KKT

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \ge 1$$

$$0 < \alpha_i < C \Leftrightarrow y_i g(x_i) = 1$$

$$\alpha_i = C \Leftrightarrow y_i g(x_i) \le 1$$

求出alpha i.

求出alpha j

这是在前面求解子问题时推导的

$(y_i y_j) \sim 1$

$$\alpha_i + y_i y_j \alpha_j = y_i \xi \longrightarrow \alpha_i = y_i \xi - y_i y_j \alpha_j$$

$$w = y_i \xi$$

$$\frac{1}{2} K_{ii} (w - s\alpha_j)^2 + \frac{1}{2} K_{jj} \alpha_j^2 + sK_{ij} (w - s\alpha_j)\alpha_j + y_i v_i (w - s\alpha_j) + y_j v_j \alpha_j - (w - s\alpha_j) - \alpha_j + c$$

求导并令导数为0

$$K_{ii}(w - s\alpha_j)(-s) + K_{jj}\alpha_j + sK_{ij}(w - 2s\alpha_j) - sy_i v_i + y_j v_j + s - 1 = 0$$

$$sy_i v_i = y_i y_j y_i v_i = y_j v_i$$

$$(K_{ii} + K_{jj} - 2K_{ij})\alpha_j - swK_{ii} - swK_{ij} - y_j v_i + y_j v_j + s - 1 = 0$$

$$(K_{ii} + K_{jj} - 2K_{ij})\alpha_j = sw(K_{ii} + K_{ij}) + y_j v_i - y_j v_j + 1 - s$$

X_i

$$(K_{ii} + K_{jj} - 2K_{ij})\alpha_j = \alpha_j^*(K_{ii} + K_{jj} - 2K_{ij}) + y_j(u_i - u_j + y_j - y_i)$$

$$\eta = K_{ii} + K_{jj} - 2K_{ij} \qquad \alpha_j^{new} = \alpha_j + \frac{y_j(E_i - E_j)}{\eta} \qquad E_i = u_i - y_i$$

这里要让

$$|E_i - E_j|$$

的绝对值最大化，

找出

$$\alpha_i \to \alpha_j$$

这就求出alpha j。

15.8 完整的算法

完整的算法
外层循环
　　选择优化变量
　　求解子问题
　　如果 已经收敛，则退出
结束循环

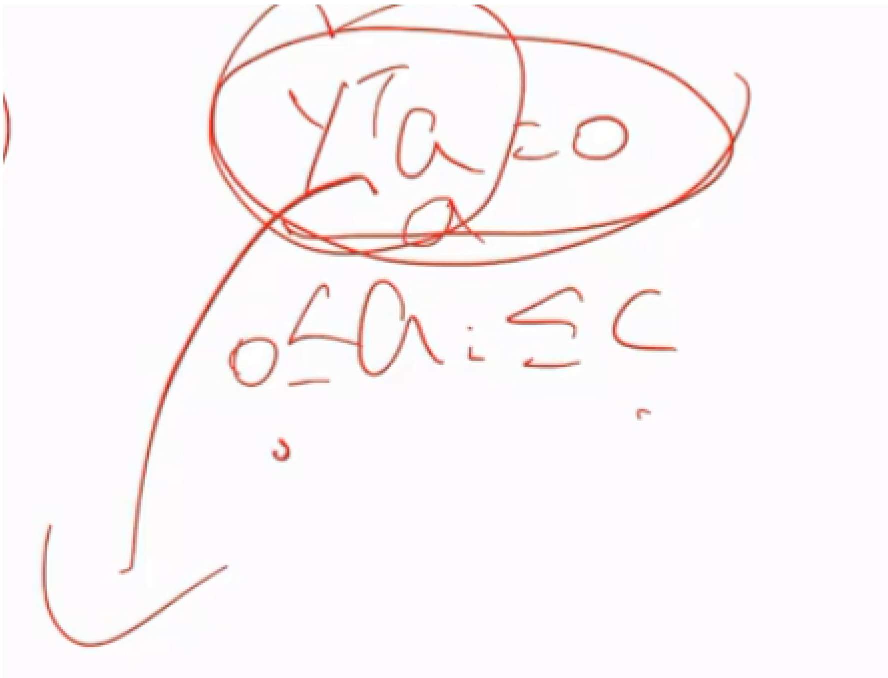
优化变量就两个alpha

完整的算法
外层循环
　　选择优化变量 → $a_i\ a_j$
　　求解子问题
　　如果 已经收敛，则退出
结束循环

15.9 实现细节问题

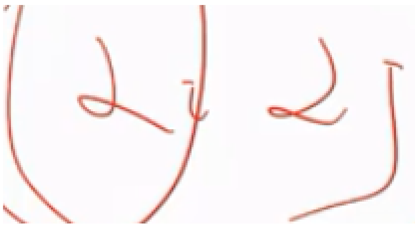
实现细节问题
初始值的设定，一般设置为全0向量
迭代终止的判定规则

- 初始值



- 迭代终止

  按照KKT条件来判断



  为了保险起见，设置最大迭代次数。

15.10 本集总结

1. why SMO?
2. SMO思想
3. 优化变量的选择
4. 完整的算法
5. 算法细节问题