第5章 支持向量机*

1. 线性可分支持向量机与硬间隔最大化
2. 线性支持向量机与软间隔最大化
3. 非线性支持向量机与核函数
4. 序列最小最优化算法

*参阅《机器学习方法》第7章

序列最小最优化算法

SMO (sequential minimal optimization)

John C. Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines" in *Advances in Neural Information Processing Systems* 11, M. S. Kearns, S. A. Solla, D. A. Cohn, eds (MIT Press, 1999), 557–63.

动机：

- 支持向量机的学习问题可以形式化为求解凸二次规划问题。这样的凸二次规划问题具有全局最优解，并且有许多最优化算法可以用于这一问题的求解；
- 但是当训练样本容量很大时，这些算法往往变得非常低效，以致无法使用。
- 所以，如何高效地实现支持向量机学习就成为一个重要的问题。

序列最小最优化算法

- SMO (sequential minimal optimization)

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$\alpha_i \geq 0, i = 1, \cdots, m$$

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \leq \alpha_i \leq C, i = 1, \cdots, m$$

对偶问题的求解

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

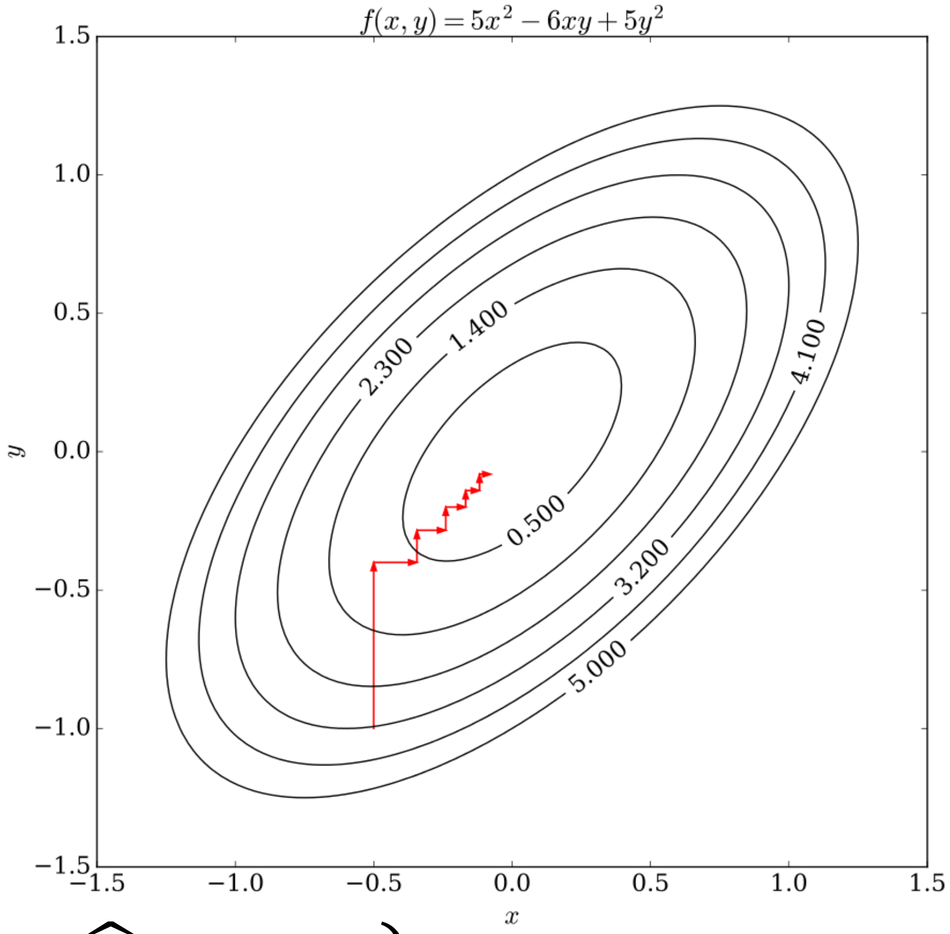
$$0 \leq \alpha_i \leq C, i = 1, \cdots, m$$

坐标下降法

- 优化问题

$$\min_{a} \theta\,(a_1, a_2, \dots, a_m)$$

- 坐标下降法



Loop until converge:{

for $i = 1, \dots, m$ {

$$a_i := \underset{\widehat{a_i}}{\operatorname{argmin}}\, \theta\,(a_1, a_2, \dots \widehat{a_i}, \dots, a_m)$$

}

}

https://en.wikipedia.org/wiki/Coordinate_descent

序列最小最优化算法

- SMO (sequential minimal optimization)

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

基本思路：

- 如果所有变量的解都满足此最优化问题的KKT条件，那么这个最优化问题的解就得到了；
- 否则，选择两个变量，固定其它变量，针对这两个变量构建一个二次规划问题，称为子问题；子问题中的两个变量：一个是违反KKT条件最严重的那个，另一个由约束条件自动确定；
- 如此，SMO算法将上述优化问题不断分解为子问题并对子问题求解（子问题可通过解析的方法求解），进而达到求解该优化问题的目的。

SMO算法

- 输入：$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，$\boldsymbol{x}_i\in\mathbb{R}^d,y_i\in\{-1,1\}$
- 输出：近似解 $\boldsymbol{\alpha}^*$

（1）令$k=0$，取初值$\boldsymbol{\alpha}^{(k)}=\mathbf{0}$；

（2）根据“启发式方法” 选取优化变量$\alpha_1^{(k)}$和$\alpha_2^{(k)}$，解析求解两个变量的最优化问题，求得最优解$\alpha_1^{(k+1)}$和$\alpha_2^{(k+1)}$，更新$\boldsymbol{\alpha}^{(k)}$为$\boldsymbol{\alpha}^{(k+1)}$；

（3）若在精度ε范围内满足停机条件

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

$$0\le\alpha_i\le C,i=1,\cdots,m$$

$$y_i g(x_i)=\begin{cases}\ge 1, & \{x_i|\alpha_i=0\}\\ =1, & \{x_i|0<\alpha_i<C\}\\ \le 1, & \{x_i|\alpha_i=C\}\end{cases}$$

$$g(x_i)=\sum_{j=1}^{m}\alpha_j y_j K(x_j,x_i)+b$$

则转（4），否则转（2）；

（4）取$\boldsymbol{\alpha}^*=\boldsymbol{\alpha}^{(k+1)}$

序列最小最优化算法

- SMO (sequential minimal optimization)

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

主要部分：

- 求解两个变量二次规划的解析方法
- 选择变量的启发式方法

$$\min_{\alpha_1\alpha_2} \quad W(\alpha_1,\alpha_2) = \frac{1}{2}K_{11}\alpha_1^2 + \frac{1}{2}K_{22}\alpha_2^2 + y_1y_2K_{12}\alpha_1\alpha_2 - (\alpha_1+\alpha_2)$$

$$+y_1\alpha_1\sum_{i=3}^{\mathrm{m}} y_i\alpha_iK_{i1} + y_2\alpha_2\sum_{i=3}^{m} y_i\alpha_iK_{i2}$$

序列最小最优化算法

◻ 两个变量二次规划的求解方法

选择两个变量α_1, α_2，其他变量 $\alpha_i (i = 3, \dots m)$ 是固定的。

构建子问题：

$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1, \alpha_2) = \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2 - (\alpha_1 + \alpha_2)$$

$$+ y_1 \alpha_1 \sum_{i=3}^{\mathrm{m}} y_i \alpha_i K_{i1} + y_2 \alpha_2 \sum_{i=3}^{m} y_i \alpha_i K_{i2}$$

$$\text{s.t.} \quad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{m} \alpha_i y_i = \varsigma, \qquad K_{ij} = K(x_i, x_j),$$

$$\underline{0 \le \alpha_i \le C, i = 1, \cdots, m} \qquad \varsigma\text{是一个常数}$$

序列最小最优化算法

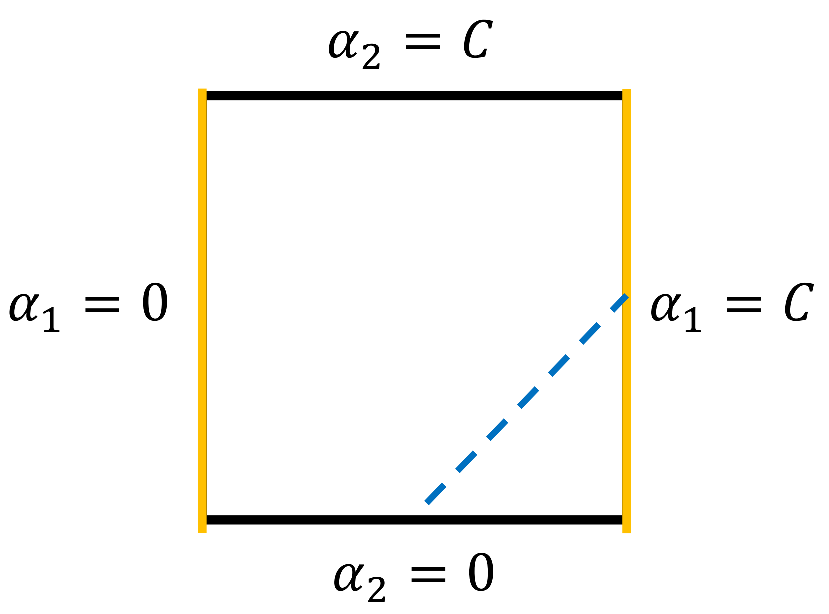
两个变量二次规划的求解方法

$$\text{s.t.}\ \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{m} \alpha_i y_i = \varsigma,$$

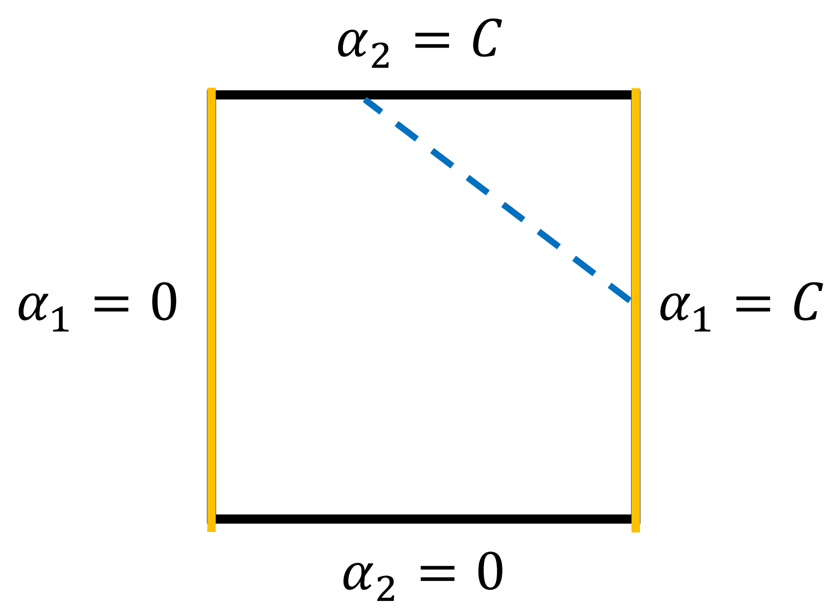
$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

"二次元"

α_2 α_1



$$y_1 \neq y_2 \Rightarrow \alpha_1 - \alpha_2 = k$$



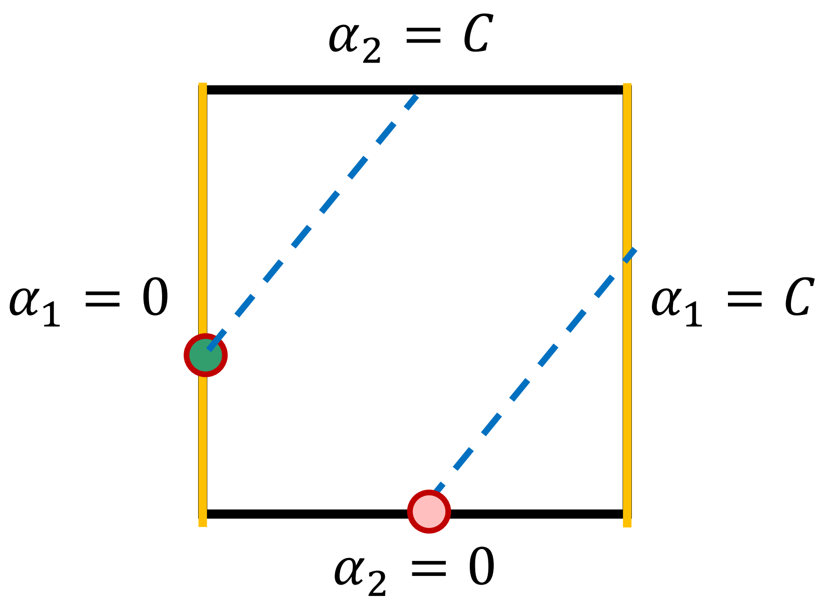
$$y_1 = y_2 \Rightarrow \alpha_1 + \alpha_2 = k$$

序列最小最优化算法

两个变量二次规划的求解方法

$$\text{s.t.} \quad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=1}^{m} \alpha_i y_i = \varsigma,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$



$$y_1 \ne y_2 \Rightarrow \alpha_1 - \alpha_2 = k$$

假设问题的初始可行解为

$\alpha_1^{old}, \alpha_2^{old}$，有

$$\alpha_1^{old} y_1 + \alpha_2^{old} y_2 = \varsigma$$

当前求得的最优解

$\alpha_1^{new}, \alpha_2^{new}$，有

$$L \le \alpha_2^{new} \le H$$

$$L = \max(0, \alpha_2^{old} - \alpha_1^{old})$$

$$\alpha_1 - \alpha_2 = k \qquad \alpha_1 = 0$$

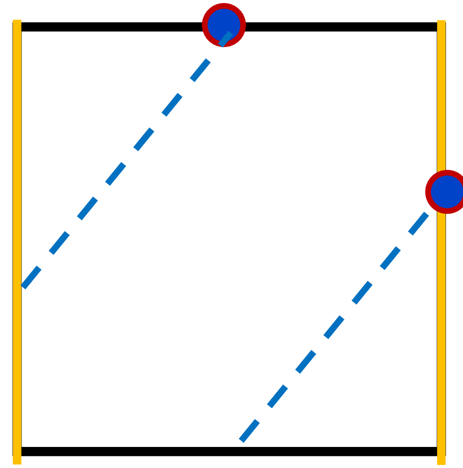
$$\alpha_2 = -k = \alpha_2^{old} - \alpha_1^{old}$$

序列最小最优化算法

两个变量二次规划的求解方法

$$\text{s.t.}\quad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=1}^{m} \alpha_i y_i = \varsigma,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$



$$y_1 \ne y_2 \Rightarrow \alpha_1 - \alpha_2 = k$$

假设问题的初始可行解为

$\alpha_1^{old}, \alpha_2^{old}$，有

$$\alpha_1^{old} y_1 + \alpha_2^{old} y_2 = \varsigma$$

当前求得的最优解

$\alpha_1^{new}, \alpha_2^{new}$，有

$$L \le \alpha_2^{new} \le H$$

$$L = \max(0, \alpha_2^{old} - \alpha_1^{old})$$

$$H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$$

$$\alpha_1 - \alpha_2 = k \qquad \alpha_1 = C$$

$$\alpha_2 = C - k = C + \alpha_2^{old} - \alpha_1^{old}$$

序列最小最优化算法

两个变量二次规划的求解方法

$$\text{s.t.}\quad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=1}^{m} \alpha_i y_i = \varsigma,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

同理可得

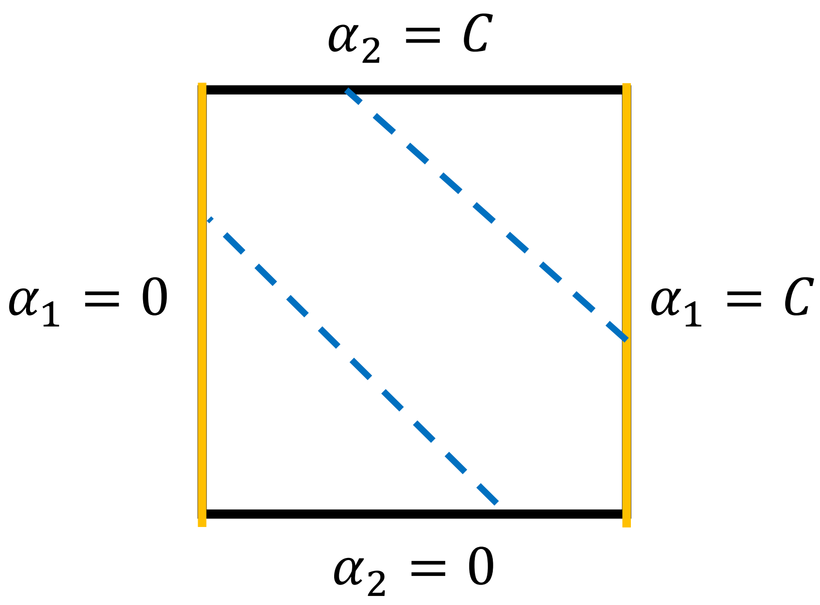
当前求得的最优解

$\alpha_1^{new}, \alpha_2^{new}$，有

$$L \le \alpha_2^{new} \le H$$

$$L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C)$$

$$H = \min(C, \alpha_2^{old} + \alpha_1^{old})$$



$$y_1 = y_2 \Rightarrow \alpha_1 + \alpha_2 = k$$

序列最小最优化算法

◻ 两个变量二次规划的求解方法

$$\min_{\alpha_1,\alpha_2} \; W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2-(\alpha_1+\alpha_2)$$
$$+y_1\alpha_1\sum_{i=3}^{m}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{m}y_i\alpha_iK_{i2}$$

引入 $\mathrm{v}_1=\sum_{i=3}^{m}y_i\alpha_iK_{i1}$ ； $\mathrm{v}_2=\sum_{i=3}^{m}y_i\alpha_iK_{i2}$

$$\min_{\alpha_1,\alpha_2} \; W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2-(\alpha_1+\alpha_2)$$
$$+y_1v_1\alpha_1+y_2v_2\alpha_2$$

序列最小最优化算法

两个变量二次规划的求解方法

$$\min_{\alpha_1,\alpha_2} \; W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2-(\alpha_1+\alpha_2)$$
$$+y_1v_1\alpha_1+y_2v_2\alpha_2$$

$\alpha_1y_1=\varsigma-\alpha_2y_2$ 且 $y_i^2=1$, 因此 $\alpha_1=(\varsigma-\alpha_2y_2)y_1$

$$\min_{\alpha_2} \; W(\alpha_2)=\frac{1}{2}K_{11}(\varsigma-\alpha_2y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+y_2K_{12}(\varsigma$$
$$-\alpha_2y_2)\alpha_2-(\varsigma-\alpha_2y_2)y_1-\alpha_2+v_1(\varsigma-\alpha_2y_2)+y_2v_2\alpha_2$$

序列最小最优化算法

◻ 两个变量二次规划的求解方法

$$\min_{\alpha_2} \quad W(\alpha_2)=\frac{1}{2}K_{11}(\varsigma-\alpha_2 y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+y_2K_{12}(\varsigma-\alpha_2 y_2)\alpha_2-(\varsigma-\alpha_2 y_2)y_1-\alpha_2+v_1(\varsigma-\alpha_2 y_2)+y_2v_2\alpha_2$$

$$\frac{\partial W(\alpha_2)}{\partial \alpha_2}=K_{11}\alpha_2+K_{22}\alpha_2-2K_{12}\alpha_2-K_{11}\,\varsigma y_2+K_{12}\,\varsigma y_2+y_1y_2-1-v_1y_2+y_2v_2$$

令 $\frac{\partial W(\alpha_2)}{\partial \alpha_2}=0$ $\qquad \alpha_2^{new,unc}=\alpha_2^{old}+\frac{y_2(E_1-E_2)}{\eta}$

$$E_i=\left(\sum_{j=1}^{m}\alpha_j y_j K(x_j,x_i)+b\right)-y_i$$

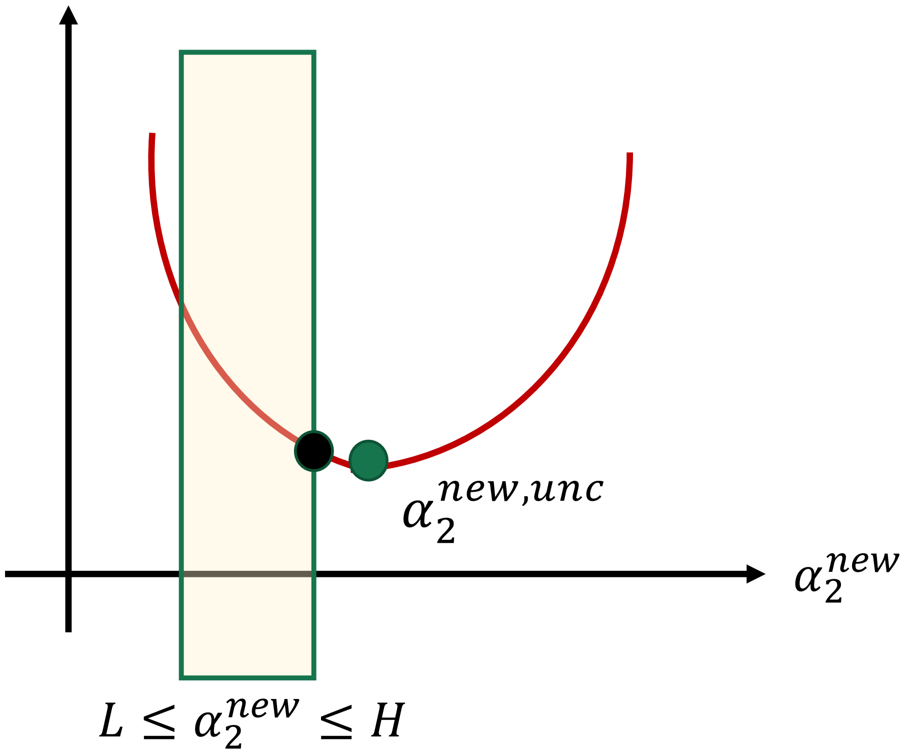
序列最小最优化算法

两个变量二次规划的求解方法

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$L \leq \alpha_2^{new} \leq H$$

$$\alpha_2^{new,unc} > H \quad \Rightarrow \quad \alpha_2^{new} = H$$

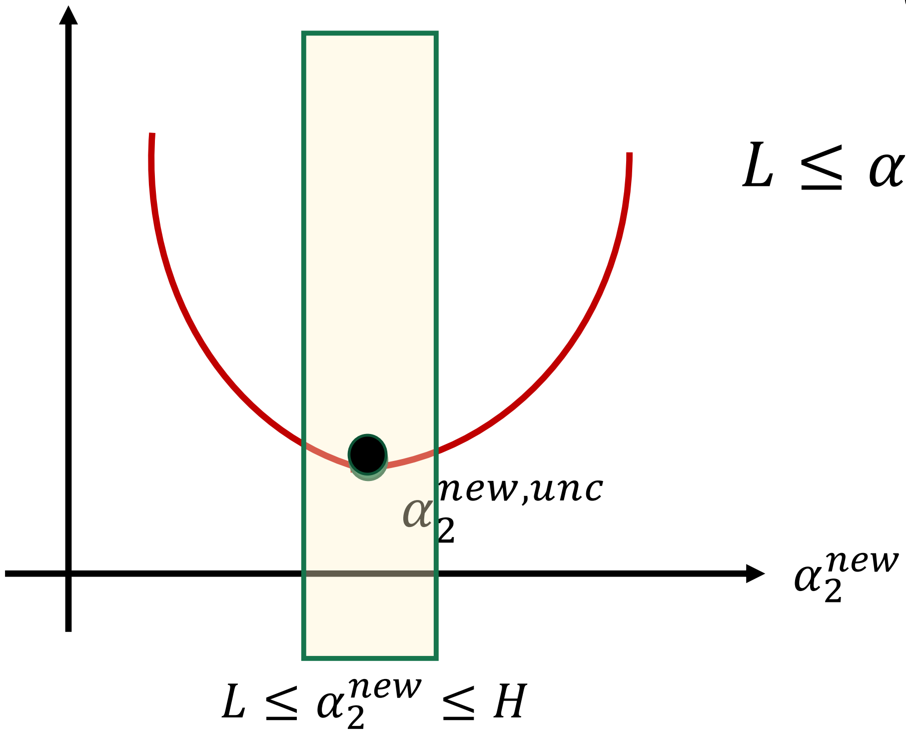


序列最小最优化算法

◻ 两个变量二次规划的求解方法

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$L \le \alpha_2^{new} \le H$$



$$\alpha_2^{new,unc} > H \quad \Rightarrow \quad \alpha_2^{new} = H$$

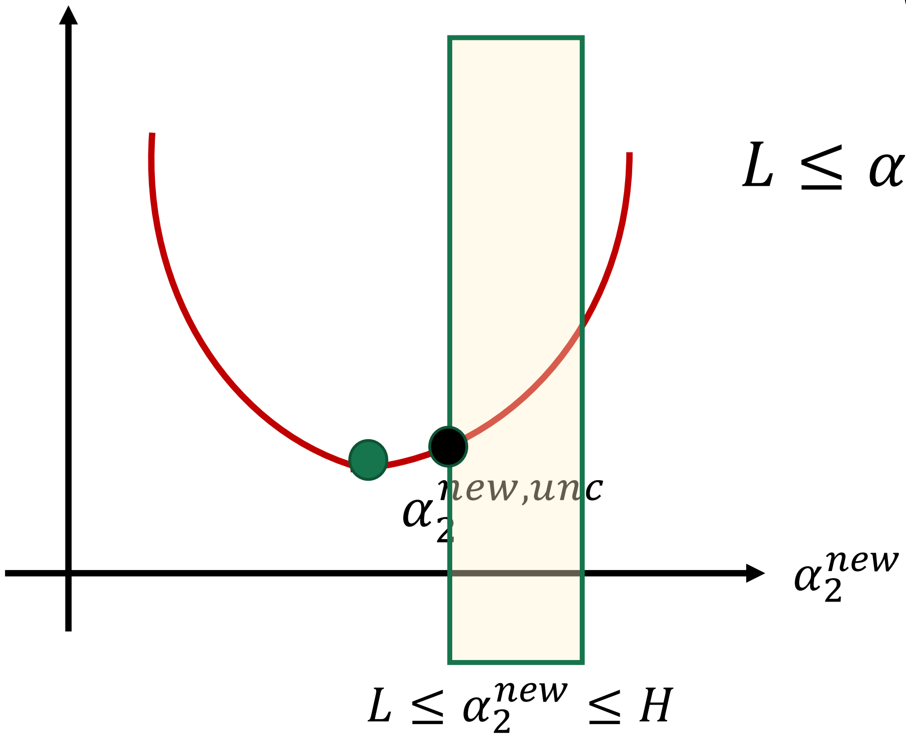
$$L \le \alpha_2^{new,unc} \le H \Rightarrow \alpha_2^{new} = \alpha_2^{new,unc}$$

序列最小最优化算法

两个变量二次规划的求解方法

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$L \le \alpha_2^{new} \le H$$



$$\alpha_2^{new,unc} > H \quad \Rightarrow \quad \alpha_2^{new} = H$$

$$L \le \alpha_2^{new,unc} \le H \Rightarrow \alpha_2^{new} = \alpha_2^{new,unc}$$

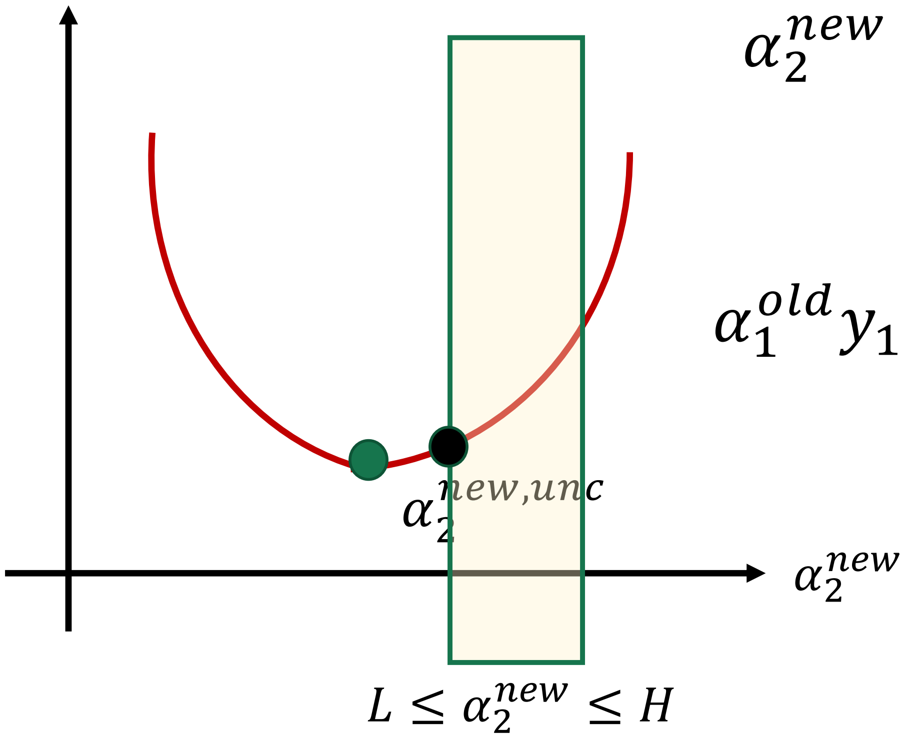
$$\alpha_2^{new,unc} < L \quad \Rightarrow \quad \alpha_2^{new} = L$$

序列最小最优化算法

■ 两个变量二次规划的求解方法

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$L \leq \alpha_2^{new} \leq H$$



$$\alpha_2^{new} = \begin{cases} H & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc} & L \leq \alpha_2^{new,unc} \leq H \\ L & \alpha_2^{new,unc} < L \end{cases}$$

$$\alpha_1^{old} y_1 + \alpha_2^{old} y_2 = \varsigma = \alpha_1^{new} y_1 + \alpha_2^{new} y_2$$

$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$$

序列最小最优化算法

- SMO (sequential minimal optimization)

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

主要部分：

- 求解两个变量二次规划的解析方法
- 选择变量的启发式方法

变量的选择方法

- 第1个变量的选择

外层循环：

- 违反KKT最严重的样本点，将其作为第1个变量

$$\alpha_i = 0 \Rightarrow y_i(\boldsymbol{\omega}^T x_i + b) \geq 1$$

$$0 < \alpha_i < C \quad \Rightarrow \quad y_i(\boldsymbol{\omega}^T x_i + b) = 1$$

$$\alpha_i = C \Rightarrow y_i(\boldsymbol{\omega}^T x_i + b) \leq 1$$

即：

- 首先遍历所有满足条件 $0 < \alpha_i < C$ 的样本点，即在间隔边界上的支持向量点
- 如果都满足，那么遍历整个训练集

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

内层循环：

- 选择的标准是希望能使目标函数有足够大的变化,即对应 $|E_1 - E_2|$最大，即E_1, E_2的符号相反，差异最大

SMO算法

- 输入：$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，$\boldsymbol{x}_i\in\mathbb{R}^d, y_i\in\{-1,1\}$
- 输出：近似解 $\boldsymbol{\alpha}^*$

（1）令$k=0$，取初值$\boldsymbol{\alpha}^{(k)}=\mathbf{0}$；

（2）根据“启发式方法” 选取优化变量$\alpha_1^{(k)}$和$\alpha_2^{(k)}$，解析求解两个变量的最优化问题，求得最优解$\alpha_1^{(k+1)}$和$\alpha_2^{(k+1)}$，更新$\boldsymbol{\alpha}^{(k)}$为$\boldsymbol{\alpha}^{(k+1)}$；

（3）若在精度ε范围内满足停机条件

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

$$0\le\alpha_i\le C, i=1,\cdots,m$$

$$y_i g(x_i)=\begin{cases}\ge 1, & \{x_i|\alpha_i=0\}\\ =1, & \{x_i|0<\alpha_i<C\}\\ \le 1, & \{x_i|\alpha_i=C\}\end{cases}$$

$$g(x_i)=\sum_{j=1}^{m}\alpha_j y_j K\left(x_j,x_i\right)+b$$

则转（4），否则转（2）；

（4）取$\boldsymbol{\alpha}^*=\boldsymbol{\alpha}^{(k+1)}$

第5章 支持向量机*

1. 线性可分支持向量机与硬间隔最大化

（数据、模型、策略、学习的对偶算法）

2. 线性支持向量机与软间隔最大化

（数据、模型、策略、学习的对偶算法）

3. 非线性支持向量机与核函数

（核技巧、常用核函数、核技巧在支持向量机中的应用）

4. 序列最小最优化算法

（二次规划求解方法、变量的选择方法、SMO算法）

*参阅《机器学习方法》第7章

思考题

已知正例点$x_1 = (1,2)^{\mathrm{T}}, x_2 = (1,3)^{\mathrm{T}}, x_3 = (3,3)^{\mathrm{T}}$，负例点$x_4 = (2,1)^{\mathrm{T}}, x_5 = (3,2)^{\mathrm{T}}, x_6 = (0,1)^{\mathrm{T}}$，试求最大间隔分离超平面和分类决策函数，并在图中画出分离超平面、间隔边界及支持向量；求解过程要求使用对偶算法和SMO方法。