SVM-作业讲解

华东师范大学 计算机科学与技术学院

张志忠 副研究员

课程内容：

- 方法回顾
- Code一些解释

回顾一下SMO 算法:

例 7.1 数据与例 2.1 相同. 已知一个如图 7.4 所示的训练数据集，其正例点是 $x_1=(3,3)^{\mathrm{T}}$， $x_2=(4,3)^{\mathrm{T}}$， 负例点是 $x_3=(1,1)^{\mathrm{T}}$， 试求最大间隔分离超平面.

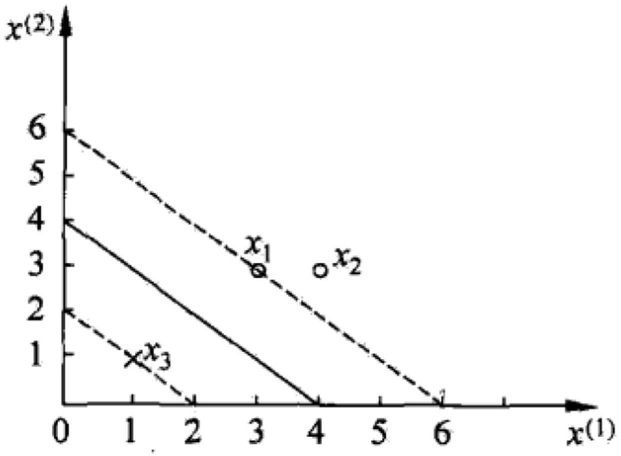


图 7.4 间隔最大分离超平面示例

原始形式

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$\text{s.t.} \quad y_i(w\cdot x_i+b)-1\geqslant 0, \quad i=1,2,\cdots,N$$

对偶形式

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i$$
$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i=0$$
$$\alpha_i\geqslant 0, \quad i=1,2,\cdots,N$$

□ 原始问题

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$\text{s.t.} \quad y_i(w\cdot x_i+b)-1\geqslant 0,\quad i=1,2,\cdots,N$$

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i$$
$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i=0$$
$$\alpha_i\geqslant 0,\quad i=1,2,\cdots,N$$

□ 拉格朗日函数

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

□ 求偏导

$$\nabla_w L(w,b,\alpha)=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0$$
$$\nabla_b L(w,b,\alpha)=\sum_{i=1}^{N}\alpha_i y_i=0$$

$$\Rightarrow \quad w=\sum_{i=1}^{N}\alpha_i y_i x_i, \qquad \sum_{i=1}^{N}\alpha_i y_i=0$$

$$\nabla_x L(x^*,\alpha^*,\beta^*)=0$$
$$\nabla_\alpha L(x^*,\alpha^*,\beta^*)=0$$
$$\nabla_\beta L(x^*,\alpha^*,\beta^*)=0$$
$$\alpha_i^* c_i(x^*)=0,\quad i=1,2,\cdots,k$$
$$c_i(x^*)\leqslant 0,\quad i=1,2,\cdots,k$$
$$\alpha_i^*\geqslant 0,\quad i=1,2,\cdots,k$$
$$h_j(x^*)=0\quad j=1,2,\cdots,l$$

KTT 条件

算法 7.2（线性可分支持向量机学习算法）

输入：线性可分训练集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in\mathcal{X}=\mathbf{R}^n$，$y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$；

输出：分离超平面和分类决策函数.

（1）构造并求解约束最优化问题

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

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

$$\alpha_i\geqslant 0,\quad i=1,2,\cdots,N$$

求得最优解 $\alpha^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_N^*)^{\mathrm{T}}$.

（2）计算

$$w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$$

并选择 α^* 的一个正分量 $\alpha_j^*>0$，计算

并选择 α^* 的一个正分量 $\alpha_j^*>0$，计算

$$b^*=y_j-\sum_{i=1}^{N}\alpha_i^* y_i(x_i\cdot x_j)$$

（3）求得分离超平面

$$w^*\cdot x+b^*=0$$

分类决策函数：

$$f(x)=\operatorname{sign}(w^*\cdot x+b^*)$$

对偶形式

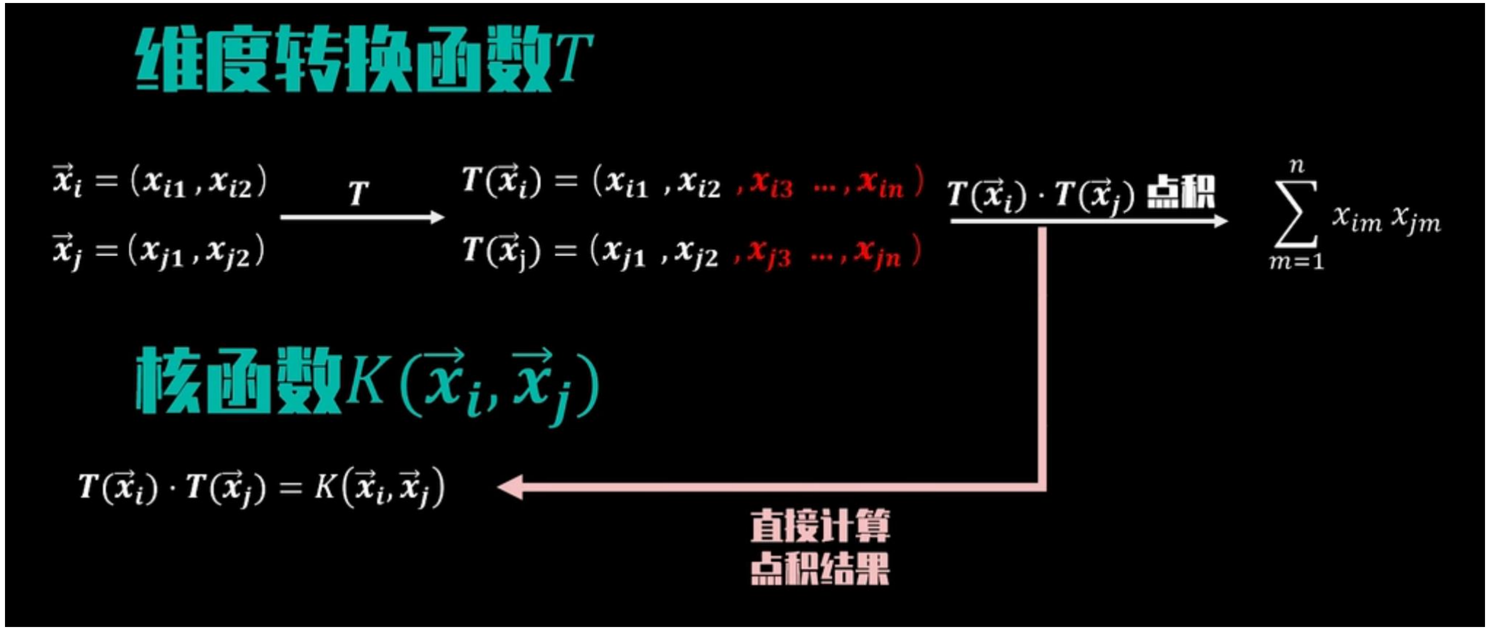
$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i \geqslant 0, \quad i = 1, 2, \cdots, N$$

多项式核： $K(x,z) = (x \cdot z + 1)^p$

高斯核： $K(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$ ➡ 分情况讨论



SMO 算法流程：

- 输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$
- $x_i\in\mathcal{X}=\mathbf{R}^n$ $y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$，精度ε
- 输出：近似解 α

（1）取初值 $\alpha^{(0)}=0$，令 $k=0$

（2）选取优化变量 $\alpha_1^{(k)},\alpha_2^{(k)}$，解析求解两个变量的最优化问题

求得最优解 $\alpha_1^{(k+1)},\alpha_2^{(k+1)}$，更新 α 为 $\alpha^{(k+1)}$；

（3）若在精度 ε 范围内满足停机条件

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

$$0\leqslant\alpha_i\leqslant C,\quad i=1,2,\cdots,N$$

$$y_i\cdot g(x_i)=\begin{cases}\geqslant 1, & \{x_i\mid\alpha_i=0\}\\ =1, & \{x_i\mid 0<\alpha_i<C\}\\ \leqslant 1, & \{x_i\mid\alpha_i=C\}\end{cases}$$

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

则转 (4)；否则令 $k=k+1$，转 (2)；

（4）取 $\hat{\alpha}=\alpha^{(k+1)}$

SMO 算法：

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

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

$$0 \leqslant \alpha_i \leqslant C, \quad i=1,2,\cdots,N$$

选取两个变量

$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2 -(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2} \tag{7.101}$$

经剪辑后 α_2 的解是

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

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

$$-(\alpha_1+\alpha_2)+y_1\alpha_1\underline{\sum_{i=3}^{N}y_i\alpha_iK_{i1}}+y_2\alpha_2\underline{\sum_{i=3}^{N}y_i\alpha_iK_{i2}} \qquad (7.101)$$

$$v_i=\sum_{j=3}^{N}\alpha_jy_jK(x_i,x_j)$$

ꕥ目标函数写成：

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

ꕥ由 $\alpha_1y_1=\varsigma-\alpha_2y_2$ 及 $y_i^2=1$

$$\alpha_1=(\varsigma-y_2\alpha_2)y_1$$

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

Minimize:

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

偏导数置为 0

将 $\eta=K_{11}+K_{22}-2K_{12}$ 代入：

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

$$g(x)=\sum_{i=1}^{N}\alpha_i y_i K(x_i,x)+b$$

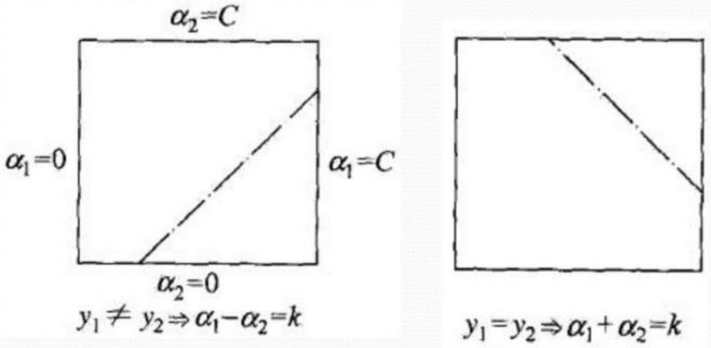
$$E_i=g(x_i)-y_i=\left(\sum_{j=1}^{N}\alpha_j y_j K(x_j,x_i)+b\right)-y_i,\quad i=1,2$$

- 根据不等式条件 α_2^{new} 的取值范围：

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

- 左图：$L = \max(0, \alpha_2^{old} - \alpha_1^{old}) \quad H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$

- 右图：$L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C) \quad H = \min(C, \alpha_2^{old} + \alpha_1^{old})$



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

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

如何选取两个变量：

1. 第 1 个变量的选择

SMO 称选择第 1 个变量的过程为外层循环．外层循环在训练样本中选取违反 KKT 条件最严重的样本点，并将其对应的变量作为第 1 个变量．具体地，检验训练样本点 (x_i, y_i) 是否满足 KKT 条件，即

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \geqslant 1 \tag{7.111}$$

先检查

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

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

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

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

2. 第 2 个变量的选择

SMO 称选择第 2 个变量的过程为内层循环．假设在外层循环中已经找到第 1 个变量 α_1，现在要在内层循环中找第 2 个变量 α_2．第 2 个变量选择的标准是希望能使 α_2 有足够大的变化．

由式 (7.106) 和式 (7.108) 可知，α_2^{new} 是依赖于 $|E_1 - E_2|$ 的，为了加快计算速度，一种简单的做法是选择 α_2，使其对应的 $|E_1 - E_2|$ 最大．因为 α_1 已定，E_1 也确定了．如果 E_1 是正的，那么选择最小的 E_i 作为 E_2；如果 E_1 是负的，那么选择最大的 E_i 作为 E_2．为了节省计算时间，将所有 E_i 值保存在一个列表中．

更新完了alpha 后，还需要更新E 和b：

$$b_1^{\text{new}} = -E_1 - y_1K_{11}(\alpha_1^{\text{new}} - \alpha_1^{\text{old}}) - y_2K_{21}(\alpha_2^{\text{new}} - \alpha_2^{\text{old}}) + b^{\text{old}} \tag{7.115}$$

同样，如果$0<\alpha_2^{\text{new}}<C$，那么，

$$b_2^{\text{new}} = -E_2 - y_1K_{12}(\alpha_1^{\text{new}} - \alpha_1^{\text{old}}) - y_2K_{22}(\alpha_2^{\text{new}} - \alpha_2^{\text{old}}) + b^{\text{old}} \tag{7.116}$$

如果$\alpha_1^{\text{new}},\alpha_2^{\text{new}}$同时满足条件$0<\alpha_i^{\text{new}}<C$，$i=1,2$，那么$b_1^{\text{new}}=b_2^{\text{new}}$．如果$\alpha_1^{\text{new}},\alpha_2^{\text{new}}$是0或者$C$，那么$b_1^{\text{new}}$和$b_2^{\text{new}}$以及它们之间的数都是符合KKT条件的阈值，这时选择它们的中点作为b^{new}．

在每次完成两个变量的优化之后，还必须更新对应的E_i值，并将它们保存在列表中．E_i值的更新要用到b^{new}值，以及所有支持向量对应的α_j：

$$E_i^{\text{new}} = \sum_S y_j\alpha_j K(x_i,x_j) + b^{\text{new}} - y_i \tag{7.117}$$

其中，S是所有支持向量x_j的集合．

SMO 算法流程：

EAST CHINA NORMAL UNIVERSITY

- 输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$
- $x_i \in \mathcal{X}=\mathbf{R}^n \quad y_i \in \mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$，精度ε
- 输出：近似解 α

（1）取初值 $\alpha^{(0)}=0$，令 $k=0$

（2）选取优化变量 $\alpha_1^{(k)},\alpha_2^{(k)}$，解析求解两个变量的最优化问题

求得最优解 $\alpha_1^{(k+1)},\alpha_2^{(k+1)}$，更新 α 为 $\alpha^{(k+1)}$；

（3）若在精度 ε 范围内满足停机条件

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

$$0\leqslant \alpha_i \leqslant C,\quad i=1,2,\cdots,N$$

$$y_i\cdot g(x_i)=\begin{cases}\geqslant 1, & \{x_i \mid \alpha_i=0\}\\ =1, & \{x_i \mid 0<\alpha_i<C\}\\ \leqslant 1, & \{x_i \mid \alpha_i=C\}\end{cases}$$

则转 (4)；否则令 $k=k+1$，转 (2)；

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

（4）取 $\hat{\alpha}=\alpha^{(k+1)}$

谢谢

Thank You