优点：泛化错误率低，计算开销不大，结果易解释

缺点：对参数调节和核函数的选择敏感，原始分类器不加修改仅适用于处理二类问题

1. SVM（原型及对偶）

$$\min_{\boldsymbol{w},b} \ \frac{1}{2}\|\boldsymbol{w}\|^2 \tag{6.6}$$

$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geqslant 1, \quad i = 1, 2, \ldots, m.$$

这就是支持向量机(Support Vector Machine, 简称 SVM)的基本型.

拉格朗日

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)\right) , \tag{6.8}$$

其中 $\boldsymbol{\alpha} = (\alpha_1; \alpha_2; \ldots; \alpha_m)$. 令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 对 $\boldsymbol{w}$ 和 b 的偏导为零可得

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i , \tag{6.9}$$

$$0 = \sum_{i=1}^{m} \alpha_i y_i . \tag{6.10}$$

将式(6.9)代入(6.8), 即可将 $L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 中的 $\boldsymbol{w}$ 和 b 消去, 再考虑式(6.10)的约束, 就得到式(6.6)的对偶问题

$$\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j \tag{6.11}$$

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

$$\alpha_i \geqslant 0 , \quad i = 1, 2, \ldots, m .$$

解出 $\boldsymbol{\alpha}$ 后, 求出 $\boldsymbol{w}$ 与 b 即可得到模型

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b \\ &= \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x} + b . \end{aligned} \tag{6.12}$$

KKT条件

$$
\begin{cases}
\alpha_i \geqslant 0\ ; \\
y_i f(\boldsymbol{x}_i) - 1 \geqslant 0\ ; \\
\alpha_i \left(y_i f(\boldsymbol{x}_i) - 1\right) = 0\ ..
\end{cases}
\tag{6.13}
$$

SVM重要性质:训练完成后，大部分的训练样本都不需保留，最终模型仅与支持向量有关。

2. 软间隔SVM

$$
\min_{\boldsymbol{w}, b, \xi_i} \ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i \tag{6.35}
$$

$$
\begin{aligned}
\text{s.t.}\ \ & y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geqslant 1 - \xi_i \\
& \xi_i \geqslant 0\ ,\ i = 1, 2, \ldots, m.
\end{aligned}
$$

$$
\begin{aligned}
L(\boldsymbol{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu}) = & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i \\
& + \sum_{i=1}^{m} \alpha_i \left(1 - \xi_i - y_i \left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right)\right) - \sum_{i=1}^{m} \mu_i \xi_i\ ,
\end{aligned}
\tag{6.36}
$$

其中 $\alpha_i \geqslant 0$, $\mu_i \geqslant 0$ 是拉格朗日乘子.

令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu})$ 对 $\boldsymbol{w}$, b, ξ_i 的偏导为零可得

$$
\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i\ , \tag{6.37}
$$

$$
0 = \sum_{i=1}^{m} \alpha_i y_i\ , \tag{6.38}
$$

$$
C = \alpha_i + \mu_i\ . \tag{6.39}
$$

将式(6.37)–(6.39)代入式(6.36)即可得到式(6.35)的对偶问题

$$
\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j \tag{6.40}
$$

$$
\begin{aligned}
\text{s.t.}\ \ & \sum_{i=1}^{m} \alpha_i y_i = 0\ , \\
& 0 \leqslant \alpha_i \leqslant C\ ,\quad i = 1, 2, \ldots, m\ .
\end{aligned}
$$

3. SVR

$$\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \tag{6.45}$$

$$\begin{aligned}
\text{s.t.} \;\; & f(\boldsymbol{x}_i) - y_i \leqslant \epsilon + \xi_i \;, \\
& y_i - f(\boldsymbol{x}_i) \leqslant \epsilon + \hat{\xi}_i \;, \\
& \xi_i \geqslant 0, \; \hat{\xi}_i \geqslant 0 \;, \; i = 1, 2, \ldots, m.
\end{aligned}$$

$$\begin{aligned}
& L(\boldsymbol{w}, b, \boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) \\
&= \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) - \sum_{i=1}^{m}\mu_i\xi_i - \sum_{i=1}^{m}\hat{\mu}_i\hat{\xi}_i \\
&+ \sum_{i=1}^{m}\alpha_i\big(f(\boldsymbol{x}_i) - y_i - \epsilon - \xi_i\big) + \sum_{i=1}^{m}\hat{\alpha}_i\big(y_i - f(\boldsymbol{x}_i) - \epsilon - \hat{\xi}_i\big) \;.
\end{aligned} \tag{6.46}$$

将式(6.7)代入, 再令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ 对 $\boldsymbol{w}$, b, ξ_i 和 $\hat{\xi}_i$ 的偏导为零可得

$$\boldsymbol{w} = \sum_{i=1}^{m}(\hat{\alpha}_i - \alpha_i)\boldsymbol{x}_i \;, \tag{6.47}$$

$$0 = \sum_{i=1}^{m}(\hat{\alpha}_i - \alpha_i) \;, \tag{6.48}$$

$$C = \alpha_i + \mu_i \;, \tag{6.49}$$

$$C = \hat{\alpha}_i + \hat{\mu}_i \;. \tag{6.50}$$

将式(6.47)–(6.50)代入式(6.46), 即可得到 SVR 的对偶问题

$$\begin{aligned}
\max_{\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}} \quad & \sum_{i=1}^{m} y_i(\hat{\alpha}_i - \alpha_i) - \epsilon(\hat{\alpha}_i + \alpha_i) \\
& - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \\
\text{s.t.} \quad & \sum_{i=1}^{m}(\hat{\alpha}_i - \alpha_i) = 0 \;, \\
& 0 \leqslant \alpha_i, \hat{\alpha}_i \leqslant C \;.
\end{aligned} \tag{6.51}$$

KKT

$$
\begin{cases}
\alpha_i(f(\boldsymbol{x}_i) - y_i - \epsilon - \xi_i) = 0 , \\
\hat{\alpha}_i(y_i - f(\boldsymbol{x}_i) - \epsilon - \hat{\xi}_i) = 0 , \\
\alpha_i\hat{\alpha}_i = 0 , \ \xi_i\hat{\xi}_i = 0 , \\
(C - \alpha_i)\xi_i = 0 , \ (C - \hat{\alpha}_i)\hat{\xi}_i = 0 .
\end{cases}
\tag{6.52}
$$

4. SMO（序列最小最优化）算法

SMO 算法是一种启发式算法，其基本思路是：如果所有变量的解都满足此最优化问题的 KKT 条件（Karush-Kuhn-Tucker conditions)，那么这个最优化问题的解就得到了．因为 KKT 条件是该最优化问题的充分必要条件．否则，选择两个变量，固定其他变量，针对这两个变量构建一个二次规划问题．这个二次规划问题关于这两个变量的解应该更接近原始二次规划问题的解，因为这会使得原始二次规划问题的目标函数值变得更小．重要的是，这时子问题可以通过解析方法求解，这样就可以大大提高整个算法的计算速度．子问题有两个变量，一个是违反 KKT 条件最严重的那一个，另一个由约束条件自动确定．如此，SMO 算法将原问题不断分解为子问题并对子问题求解，进而达到求解原问题的目的．

注意，子问题的两个变量中只有一个是自由变量．假设 α_1，α_2 为两个变量，$\alpha_3,\alpha_4,\cdots,\alpha_N$ 固定，那么由等式约束 (7.99) 可知

$$
\alpha_1 = -y_1\sum_{i=2}^{N}\alpha_i y_i
$$

如果 α_2 确定，那么 α_1 也随之确定．所以子问题中同时更新两个变量．

整个 SMO 算法包括两个部分：求解两个变量二次规划的解析方法和选择变量的启发式方法．

ps:李航《统计学习方法》7.4节，讲解了如何选择和计算两个变量