支持向量机SVM分类与回归算法推导过程

Friday, April 26, 2024   2:04 PM

支持向量机是一种代理模型，并支持分类与回归两种算法。

由期望风险为：

$$R(w) = \int L(y, f(x,w))\, dP(x,y)$$

(1) 经验风险 $\longrightarrow R_{emp}(w) = \frac{1}{l}\sum_{i=1}^{l} L(y_i, f(x_i, w))$

一般地，R_{emp}最小化往往可能导致过学习，而最小二乘和神经网络一般是以R_{emp}最小为准则提出的。

(2) 结构风险：

$$R(w) \le R_{emp}(w) + \Omega\left(\frac{l}{h}\right)$$

参考支持向量机部分，我们可通过最小化 $\frac{1}{2}\|w\|^2$ 来最大化平面距离 $\gamma = \frac{2}{\|w\|}$

引入松弛变量 ξ_i 之后，引入Lagrange函数为：

$$L(w,b,\xi,\alpha,\gamma) = \boxed{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i} - \sum_{i=1}^{l}\alpha_i\left[y_i(w\cdot\varphi(x_i)+b)-1+\xi_i\right] - \sum_{i=1}^{l}\gamma_i\xi_i$$

其中 w, b, ξ_i 为优化变量，$\alpha_i \ge 0$，$\gamma_i \ge 0$，为Lagrange乘子

（约束函数是 $y_i(w\cdot\varphi(x_i)+b) \ge 1-\xi_i$　　$\xi_i \ge 0$）　求

我们是求解 α, γ 在使L最大情况下的L最小值

$$\min_{x}\max_{\alpha,\gamma\ge 0} L(w,b,\xi,\alpha,\gamma) \xrightarrow{\text{对偶}} \max_{\alpha,\gamma\ge 0}\min L(w,b,\xi,\alpha,\gamma)$$

此时：利用KKT条件，有：$f(x)+\sum_{i=1}^{m}\alpha_i g_i(x)+\sum_{j=1}^{k}\beta_j h_j(x)$，$g_i(x)\le 0$，$h_i(x)=0$ 时：

$$\begin{cases} \frac{\partial L(x,\alpha^*,\beta^*)}{\partial x} = 0 \\ \alpha_i^* g_i(x) = 0 \\ h_j(x^*)=0,\ g_i(x^*)<0 \\ \alpha_i(x^*) \ge 0 \end{cases}$$

$\Longrightarrow \frac{\partial L}{\partial w} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi_i} = 0$

代入：

由第一式，得：

① $\|w\| - \sum_{i=1}^{l}\alpha_i y_i \varphi(x_i) = 0$　(1)

② $-\sum_{i=1}^{l}\alpha_i y_i\ 0, \longrightarrow \sum_{i=1}^{l}\alpha_i y_i = 0$　(2)

③ $C - \alpha_i - \gamma_i = 0 \longrightarrow \gamma_i = C-\alpha_i$　(3)

另外：$\alpha_i \ge 0$，$\gamma_i \ge 0$（初始条件限制）

代入：有：

$$L(w,b,\xi,\alpha,\gamma) = \frac{1}{2}\sum_{i=1}^{l}\alpha_i y_i \varphi(x_i)\cdot\sum_{j=1}^{l}\alpha_j y_j \varphi(x_j) + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}(C-\alpha_i)\xi_i - \sum_{i=1}^{l}\alpha_i y_i\left(\sum_{j=1}^{l}\alpha_j y_j\right)\cdot\varphi(x_i) - \cancel{b\sum_{i=1}^{l}\alpha_i y_i} - \sum_{i=1}^{l}\alpha_i(\xi_i-1)$$

$$= -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \varphi(x_i) + \cancel{\sum_{i=1}^{l}\alpha_i\xi_i} + \sum_{i=1}^{l}\alpha_i(1-\cancel{\xi_i})$$

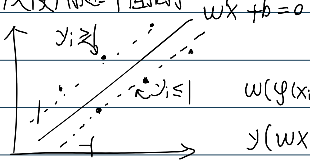
$$= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \varphi(x_i)$$

问题转化：

$$\Rightarrow \boxed{\max\left(\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \varphi(x_i) - \sum_{i=1}^{l}\alpha_i\right),\ s.t.\ \sum_{i=1}^{l}\alpha_i y_i = 0},\quad \alpha_i \in [0, C]$$

需要指出，其中函数 $\varphi(x_i)$ 用于将原空间映射到特征空间。

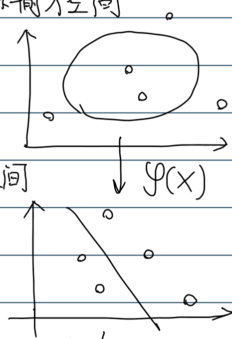
对于仅使用超平面时



$w(\varphi(x_i)+b)=wx_i+b$

$y(wx_i+b)\geq1$

$\Rightarrow$



因而我们用核函数代替 $\varphi(x)$，
则有：

$$\boxed{\max\left(\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j k(x_i,x_j)-\sum_{i=1}^{l}\alpha_i\right)} \quad \left(\begin{array}{l}\sum_{i=1}^{l}\alpha_i y_i=0\\ 0\leq\alpha_i\leq C\end{array}\right)$$

后面两项都必为0

互补松弛条件

显然：$\alpha\in[0,C]$ 而由 $\alpha_i^* g_i(x)=0$，则 $\boxed{\alpha_i(y_i(w\cdot\varphi(x_i)+b)-1+\xi_i)=0}$，(1)

若 $\alpha_i\neq0$ 时，显：$y_i(w\varphi(x_i)+b)-1+\xi_i=0$，

又：由互补松弛条件(2)有：$\gamma_i\xi_i=0$，代入 $\gamma_i=C-\alpha_i$

此时：若 $\alpha_i\in(0,C)$ 则显有 $\xi_i=0$，则对应式为

$$\boxed{\alpha_i(y_i(w\varphi(x_i)+b))=1} \quad (\alpha\in(0,C))$$

通过此式可以经过任意一个 α_i, y_i 式子求解出 b 值

而我们往往通过计算所有 b 值并取平均获取 b。

而显然二分类可用分类函数由 $w\varphi(x_i)+b$ 获取，代入有：

$$\boxed{g_{svm}(x)=sgn\left(\sum_{i=1}^{l}\alpha_i y_i k(x,x_i)+b\right)}$$ 为分类预测函数

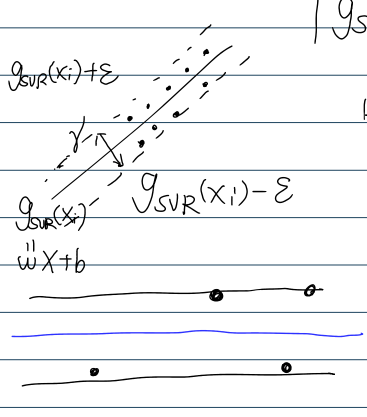
支持向量机回归优化公式推导：

我们考虑训练样本集：

$D=\{(x_1,y_1)(x_2,y_2)\cdots(x_l,y_l)\}$，而假设对于该样本集

存在一个线性近似使得：

$$|g_{SVR}(x_i)-y_i|\leq\varepsilon,\quad i=1,2,\cdots l$$



由：直线距离公式：由：$y_1=w^Tx+b$，$y_2=w^Tx+b+\varepsilon$

$$\gamma=\frac{2|\varepsilon|}{\sqrt{\|w\|^2+1}} \quad 对应地：d_i=\frac{|w^Tx+b-y_i|}{\sqrt{\|w\|^2+1}}\leq\frac{|\varepsilon|}{\sqrt{\|w\|^2+1}},\quad i=1,2,\cdots l。$$

按照支持向量理论所建立的模型即为最大化样本点之间的距离的，而显然同时满足了 $\delta\leq\varepsilon$ 这一需求。

回归问题变为：$\min\frac{1}{2}\|w^2\|$ $\Rightarrow$

回归问题变为：$\min \frac{1}{2}\|w^2\|$

st：$|g_{SVR}(x_i)-y_i| = |w^T X + b - y_i| \leqslant \varepsilon$ （$\forall i=1,2,\cdots l$ 均成立）

另外：考虑到上式不能严格满足，可引入变量 $\xi_i \geqslant 0$，$\xi_i^* \geqslant 0$，同时考虑

仍然使用拉格朗日乘数法求解：　其中 C：正则化常数

$$\min\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i+\xi_i^*)\right]$$

需说明对于回归容差线内的点：ξ_i，ξ_i^* 均为 0 ☆

约束条件松弛为：　使用不敏感损失

$$\begin{cases} g_{SVR}(x_i) - y_i - \varepsilon - \xi_i \leqslant 0 \\ y_i - g_{SVR}(x_i) - \varepsilon - \xi_i^* \leqslant 0 \end{cases} \Longrightarrow \text{分别表示两侧松弛度}$$

$$|y - g_{SVR}(x)| = \begin{cases} 0 & \cdots < \varepsilon \\ |y - g_{SVR}| - \varepsilon & \text{other} \end{cases}$$

其中：$\xi_i^*, \xi_i \geqslant 0$　$i = 1, 2, \cdots l$.

由于求解最小值，原问题为：

$$\min_{\substack{w,b,\mu\geqslant 0\\ \xi_1,\xi_2,\ \mu^*\geqslant 0\\ \alpha\geqslant 0\\ \alpha^*\geqslant 0}} \max L(w, b, \xi_1, \xi_2, \mu, \mu^*) \overset{\text{对偶}}{=} \max_{\substack{\mu\geqslant 0\\ \mu^*\geqslant 0\\ \alpha\geqslant 0\\ \alpha^*\geqslant 0}} \min_{\substack{w\\ b\\ \xi_1\\ \xi_2}} L(w, b, \xi_1, \xi_2, \mu, \mu^*)$$

为求解极值点，构造拉格朗日天条件极值　求导：

$$w - \sum\alpha_i x_i + \sum\alpha_i^* x_i = 0$$

$$\therefore w = \sum(\alpha_i - \alpha_i^*) x_i$$

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}\mu_i\xi_i - \sum_{i=1}^{l}\mu_i^*\xi_i^*$$

$$+\sum_{i=1}^{l}\alpha_i\left(g_{SVR}(x_i) - y_i - \varepsilon - \xi_i\right) + \sum_{i=1}^{l}\alpha_i^*\left(y_i - g_{SVR}(x_i) - \varepsilon - \xi_i^*\right) \text{ 求最小}$$

显然：$\mu_i\xi_i = 0$，$\mu_i^*\xi_i^* = 0$ 时最小值　其中 $g_{SVR}(x_i) = wx + b$

$\frac{1}{2}(\alpha_i - \alpha_j^*)\cdot x_i x_j - (\alpha_i - \alpha_i^*) w x_i$

①：极值点

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial C} = \frac{\partial L}{\partial \xi_i} = \frac{\partial L}{\partial \xi_i^*} = 0$$

直接求解，代入有：

$$w = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)x_i \ ①, \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \ ② \qquad \mu_i^* = C - \alpha_i^*$$

$$C - \mu_i - \alpha_i = 0 \ ③ \quad C - \mu_i^* - \alpha_i^* = 0 \ ④ \ \text{代入得：} \mu_i = C - \alpha_i$$

$$L = -\frac{1}{2}\sum_{i=1}^{l}(\alpha_i^* - \alpha_i)x_i\sum_{j=1}^{l}(\alpha_j^* - \alpha_j)x_j + \sum_{i=1}^{l}\cancel{\left[-\alpha_i\xi_i - \alpha_i^*\xi_i^* + \alpha_i\xi_i + \alpha_i^*\xi_i^*\right]}$$

$$+\sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i - \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)\varepsilon$$

$=$

$$+\sum_{i=1}(\alpha_i^*-\alpha_i)y_i-\sum_{i=1}(\alpha_i+\alpha_i^*)\varepsilon$$

$$=\underbrace{\max_{\alpha_i,\alpha_i^*\geq 0}\min_{\omega,b} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)x_ix_j+\sum_{i=1}^{l}\left[(\alpha_i^*-\alpha_i)y_i-(\alpha_i+\alpha_i^*)\varepsilon\right]}$$

$$s.t.\ \sum_{i=1}^{l}(\alpha_i^*-\alpha_i)=0;\quad 0\leq\alpha_i,\alpha_i^*\leq C\ (\text{由于}\mu_i\text{限制}).$$

而对于非线性问题，我们只需用核函数将 x 映射到高维空间：

$$\max_{\alpha_i,\alpha_i^*\geq 0}\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)y_i-(\alpha_i+\alpha_i^*)\varepsilon-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)\ x_ix_j\ K(x_i,x_j)$$

另外：利用松弛条件，有：

$$\to \mu_i=C-\alpha_i$$

$$\sum_{i=1}^{l}\alpha_i(g_{SVR}(x_i)-y_i-\varepsilon-\xi_i)=0,\quad \mu_i\xi_i=0$$

$$\sum_{i=1}^{l}\alpha_i^*(y_i-g_{SVR}(x_i)-\varepsilon-\xi_i^*)=0,\quad \mu_i^*\xi_i^*=0$$

又：$g_{SVR}=\omega\varphi(x)+b=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)k(x_i,x_j)+b$ ☆

因而：

当 $\alpha_i\in(0,C)$ 时，显有 $\xi_i=0$，$\alpha_i\neq 0$，则得：

$$\begin{cases} g_{SVR}(x_i)-y_i-\varepsilon=0, & \alpha_j\in(0,C)\\ y_i-g_{SVR}(x_i)-\varepsilon=0 & \alpha_j^*\in(0,C)\end{cases}$$

代入：
$$\begin{cases} b=y_i+\varepsilon-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)k(x_i,X) & \alpha_j\in(0,C)\\ b=y_i-\varepsilon-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)k(x_i,X) & \alpha_j^*\in(0,C)\end{cases}$$

我们只需求解最大化式子的 α_i,α_i^*，然后代入☆

$$g_{SVR}=\omega^Tx+b=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)k(x_i,x_j)+b \text{ 即可.}$$