Support Vector Machine

1 Support Vector Machine

1.1 Background

1.1.1 Optimized Problem

Constrained optimized problem:

$$\begin{aligned} &\min_{x\in\mathbf{R}^p} f(x) \\ &s.t.\ \ m_i(x) \leq 0, i=0,1,\cdots,M \\ &\qquad n_j(x) = 0, j=1,2,\cdots,N \end{aligned} \tag{1}$$

Define the Lagrange function:

$$L(x,\lambda,\eta) = f(x) + \sum_{i=1}^{M}\lambda_i m_i(x) + \sum_{j=1}^{N}\eta_j n_j(x) \tag{2}$$

The question in Eq. 1 is:

$$\min_{x\in\mathbf{R}^p}\max_{\lambda,\eta} L(x,\lambda,\eta),\ s.t.\ \lambda_i \geq 0 \tag{3}$$

1.2 Dual Problem

The dual problem of Eq. 4 is:

$$\max_{\lambda,\eta}\min_{x\in\mathbf{R}^p} L(x,\lambda,\eta),\ s.t.\ \lambda_i \geq 0 \tag{4}$$

The dual problem is the maximum problem of (λ,η).

1.3 Karush-Kuhn-Tucker (KKT) Condition

$$\begin{aligned} &\min_{x\in\mathbf{R}^p} f(x) \\ &s.t.\ \ h_i(x) = 0,\ \ i=1,2,\cdots,m \\ &\qquad g_j(x) \leq 0,\ \ j=1,2,\cdots,n \end{aligned} \tag{5}$$

Introduce the Lagrange multiplier $\lambda = (\lambda_1,\lambda_1,\cdots,\lambda_m)^T$ and $\eta = (\eta_1,\eta_2,\cdots,\eta_j)^T$. Therefore the Lagrange function is:

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m}\lambda_i h_i(x) + \sum_{i=1}^{m}\eta_j g_j(x) \tag{6}$$

The KKT condition is:

$$\begin{cases} g_j(x) \leq 0, \\ \eta_j \geq 0, \\ \eta_j g_j(x) = 0 \end{cases} \tag{7}$$

1.4 Hard Margin SVM

Let the hyper plate is

$$w^T x + b = 0 \tag{8}$$

Therefore the maximum margin is:

$$\begin{aligned} &\arg\max_{w,b}\left[\min_i \frac{|w^T x + b|}{||W||}\right] \\ &s.t.\ \ y_i(w^T x_i + b) > 0 \\ &\therefore \\ &\arg\max_{w,b}\left[\min_i \frac{y_i(w^T x + b)}{||W||}\right] \\ &s.t.\ \ y_i(w^T x_i + b) > 0 \end{aligned} \tag{9}$$

$$\begin{aligned}
&\arg\min_{w,b}\left[\frac{1}{2}w^T w\right] \\
&s.t.\quad \min_i y_i(w^T x_i + b) = 1 \\
&\therefore \\
&\arg\min_{w,b}\left[\frac{1}{2}w^T w\right] \\
&s.t.\quad \min_i y_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \cdots, N
\end{aligned} \tag{10}$$

Therefore, the Lagrange function is:

$$\begin{aligned}
L(w, b, \alpha) &= \frac{1}{2}w^T w + \sum_{i=0}^{m} \alpha_i(1 - y_i(w^T x_i + b)) \\
&= \frac{1}{2}w^T w + \sum_{i=0}^{m} \alpha_i - w^T \sum_{i=0}^{m} \alpha_i y_i x_i - b\sum_{i=0}^{m} \alpha_i y_i \\
\therefore \min_{w,b} \max_{\alpha} L(w, b, \alpha) &= \frac{1}{2}w^T w + \sum_{i=0}^{m} \alpha_i - w^T \sum_{i=0}^{m} \alpha_i y_i x_i - b\sum_{i=0}^{m} \alpha_i y_i
\end{aligned} \tag{11}$$

The dual problem of Eq. 11 is:

$$\max_{\alpha} \min_{w,b} L(w, b, \alpha) = \frac{1}{2}w^T w + \sum_{i=0}^{m} \alpha_i - w^T \sum_{i=0}^{m} \alpha_i y_i x_i - b\sum_{i=0}^{m} \alpha_i y_i \tag{12}$$

Compute the partial derivation for (w, b),

$$\begin{aligned}
\because\quad & \frac{\partial L}{\partial w} = 0 \\
\therefore\quad & w - \sum_{i=0}^{m} \alpha_i y_i x_i = 0 \\
& w^* = \sum_{i=0}^{m} \alpha_i y_i x_i \\
\because\quad & \frac{\partial L}{\partial b} = 0 \\
\therefore\quad & \sum_{i=0}^{m} \alpha_i y_i = 0
\end{aligned} \tag{13}$$

Therefore,

$$\begin{cases}
\min_{w,b} L(w, b, \alpha) \\
\frac{\partial L}{\partial w} \Rightarrow w = \sum_{i=0}^{m} \alpha_i y_i x_i \\
\frac{\partial L}{\partial w} \Rightarrow \sum_{i=0}^{m} \alpha_i y_i = 0
\end{cases} \tag{14}$$

So. bring Eq. 14 to Eq. 11,

$$\begin{aligned}
L(w, b, \alpha) &= \frac{1}{2}w^T \sum_{i=0}^{m} \alpha_i y_i x_i + \sum_{i=0}^{m} \alpha_i - w^T \sum_{i=0}^{m} \alpha_i y_i x_i \\
&= \sum_{i=0}^{m} \alpha_i - \frac{1}{2}w^T \sum_{i=0}^{m} \alpha_i y_i x_i \\
&= \sum_{i=0}^{m} \alpha_i - \frac{1}{2}\sum_{j=0}^{m} \alpha_j y_j x_j^T \sum_{i=0}^{m} \alpha_i y_i x_i \\
\max_{\alpha} L(\alpha) &= \sum_{i=0}^{m} \alpha_i - \frac{1}{2}\sum_{i=0}^{m}\sum_{j=0}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j
\end{aligned} \tag{15}$$

Therefore,

$$\begin{cases} \min\limits_{\alpha} \frac{1}{2}\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum\limits_{i=0}^{m}\alpha_i \\ s.t.\ \ \alpha_i \geq 0 \\ \qquad \sum\limits_{i=0}^{n}\alpha_i y_i = 0 \end{cases} \tag{16}$$

The KKT condition is,

$$\begin{cases} \frac{\partial L}{\partial w} = 0, \\ \frac{\partial L}{\partial b} = 0, \\ \alpha_k(1 - y_k(w^T x_k + b)) = 0, \quad \text{slackness complementary} \\ \alpha_i \geq 0, \\ 1 - y_i(w^T x_i + b) \leq 0 \end{cases} \tag{17}$$

The slackness complementary requires (α_k, x_k, y_k) meet the optimal condition. Therefore, bring $\hat{w} = \sum\limits_{i=0}^{m}\alpha_i x_i y_i$ to the slackness complementary. We can get.

$$\begin{aligned} &\hat{b} = y_k - w^T x_k = y_k - \sum_{i=0}^{m}\alpha_i y_i x_i^T x_k \\ &\exists k, 1 - y_k(w^T x_k + b) = 0 \end{aligned} \tag{18}$$

In the end, the solution of SVM is,

$$\begin{cases} \hat{w} = \sum\limits_{i=0}^{m}\alpha_i x_i y_i \\ \hat{b} = y_k - \sum\limits_{i=0}^{m}\alpha_i y_i x_i^T x_k \\ \exists k, 1 - y_k(w^T x_k + b) = 0 \end{cases} \tag{19}$$

1.5 Soft Margin SVM

Hard margin SVM makes the classification results correct for all the data in the dataset. Soft margin SVM adds the error possibility in the classification results. The number of incorrect results are,

$$\text{error} = \sum_{i=0}^{N}\mathbf{I}(y_i(w^T x_i + b) < 1) \tag{20}$$

where $\mathbf{I}$ is the indicator function. Rewrite this function as Hinge Function.

$$\text{error} = \sum_{i=0}^{m}\max(0, 1 - y_i(w^T x_i + b)) \tag{21}$$

So, the soft margin svm is

$$\begin{aligned} &\begin{cases} \min\limits_{w,b} \frac{1}{2}w^T w + C\sum\limits_{i=0}^{m}\max(0, 1 - y_i(w^T x_i + b)) \\ s.t.\ \ y_i(w^T x_i + b) \geq 1 - \xi_i \end{cases} \\ &\begin{cases} \min\limits_{w,b} \frac{1}{2}w^T w + C\sum\limits_{i=0}^{m}\xi_i \\ s.t.\ \ y_i(w^T x_i + b) \geq 1 - \xi_i \\ \qquad \xi_i \geq 0 \end{cases} \end{aligned} \tag{22}$$

Therefore the Lagrange function is,

$$L(w, b, \alpha, \xi, \mu) = \frac{1}{2}w^T w + C\sum_{i=0}^{m}\xi_i + \sum_{i=0}^{m}\alpha_i(1 - \xi_i - y_i(w^T x_i + b)) - \sum_{i=0}^{m}\mu_i\xi_i \tag{23}$$

The original problem is,

$$\min_{w,b,\xi}\max_{\alpha,\mu} L(w, b, \alpha, \xi, \mu) \tag{24}$$

The dual problem is,

$$\max_{\alpha,\mu}\min_{w,b,\xi} L(w, b, \alpha, \xi, \mu) \tag{25}$$

To get the $\min_{w,b} L$, compute the partial derivative of L for (w, b, ξ),

$$\begin{aligned} \because \quad & \frac{\partial L}{\partial w} = 0 \\ \therefore \quad & w - \sum_{i=0}^{m} \alpha_i x_i y_i = 0 \\ & w = \sum_{i=0}^{m} \alpha_i x_i y_i \\ \because \quad & \frac{\partial L}{\partial b} = 0 \\ \therefore \quad & \sum_{i=0}^{m} \alpha_i y_i = 0 \\ \because \quad & \frac{\partial L}{\partial \xi} = 0 \\ \therefore \quad & C = \alpha_i + \mu_i \end{aligned} \tag{26}$$

Therefore,

$$\begin{cases} \max\limits_{\alpha} L = \sum\limits_{i=0}^{m} \alpha_i - \frac{1}{2} \sum\limits_{i=0}^{m} \sum\limits_{j=0}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ s.t. \;\; \sum\limits_{i=0}^{m} \alpha_i y_i = 0 \\ \qquad 0 \leq \alpha_i \leq C \end{cases} \tag{27}$$

The KKT condition is,

$$\begin{cases} \alpha_i \geq 0, \;\; \mu_i \geq 0 \\ y_i f(x_i) - 1 + \xi_i \geq 0 \\ \alpha_i (y_k f(x_k) - 1 + \xi_k) = 0, \exists k, \;\; y_k f(x_k) - 1 + \xi_k = 0 \\ \xi_i \geq 0, \;\; \mu_i \xi_i = 0 \end{cases} \tag{28}$$

1.6 Support Vector Regression (SVR)

The SVR problem is,

$$\begin{aligned} & \min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \iota_i (f(x_i) - y_i) \\ & \iota_i = \begin{cases} 0, \;\; \text{if } |z| \leq \epsilon \\ |z| - \epsilon, \text{otherwise} \\ \epsilon = f(x_i) - y_i \end{cases} \end{aligned} \tag{29}$$

Introduce the slack variable ξ_i and $\hat{\xi}_i$, which are

$$\begin{cases} \xi_i = f(x_i) - y_i - \epsilon \\ \hat{\xi}_i = y_i - f(x_i) - \epsilon \end{cases} \tag{30}$$

Therefore the SVR can be written as:

$$\begin{aligned} & \min_{w,b,\xi,\hat{\xi}} \frac{1}{2} w^T w + C \sum_{i=1}^{m} (\xi_i - \hat{\xi}_i) \\ & s.t. \;\; f(x_i) - y_i \leq \epsilon + \xi_i \\ & \qquad y_i - f(x_i) \leq \epsilon + \hat{\xi}_i \\ & \qquad \xi_i \geq 0, \hat{\xi}_i \geq 0 \end{aligned} \tag{31}$$

The Lagrange function is,

$$\begin{aligned} & L(w, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu}) \\ & = \frac{1}{2} w^T w + C \sum_{i=1}^{m} (\xi_i + \hat{\xi}_i) - \sum_{i=1}^{m} \mu_i \xi_i - \sum_{i=1}^{m} \hat{\mu}_i \hat{\xi}_i + \sum_{i=1}^{m} \alpha_i (f(x_i) - y_i - \epsilon - \xi_i) + \sum_{i=1}^{m} \hat{\alpha}_i (f(x_i) - y_i - \epsilon - \hat{\xi}_i) \end{aligned} \tag{32}$$