1 SVM

1. Choose the decision boundary with the largest margin $\rightarrow$ maximum margin solution

2. Location of decision boundary is determined by a subset of the data point $\rightarrow$ support vectors

3. How to determine a plane $\rightarrow$ decision boundary, when we know the normal vector $w^T = (w_1, w_2, ..., w_n)$ and one support vector $x^* = (x_1^*, x_2^*, ..., x_n^*)$. The plane is

$$w_1(x_1 - x_1^*) + w_2(x_2 - x_2^*) + ... + w_n(x_n - x_n^*) = 0 \tag{1}$$

4. Hinge Loss: label $y \in \{-1, 1\}$, $\eta = f(x) = w^T x + w_0$.

$$L_{hinge} = max(0, 1 - y \times \eta) \tag{2}$$

1.1 derivation

In feature space the hyperplane $f(x) = w^T \phi(x) + b = 0$, then the distance of data point x^* to hyperplane is $\frac{f(x^*)}{||w||}$. label $y \in \{-1, 1\}$. Then the objective function is:

$$\arg\max_{w,b}\{\min_n \frac{y_n f(x_n)}{||w||}\} = \arg\max_{w,b}\{\min_n \frac{y_n(w^T\phi(x_n) + b)}{||w||}\} \tag{3}$$

We choose $\alpha > 0$ to let

$$\begin{aligned} w &\rightarrow \alpha w \\ b &\rightarrow \alpha b \end{aligned} \tag{4}$$

$$\min_n y_n(\alpha w^T \phi(x_n) + \alpha b) = 1 \tag{5}$$

Then the problem would be:

$$\min_{w,b} ||w|| = \min_{w,b} \frac{1}{2}||w||^2 \tag{6}$$

subject to

$$y_n(w^T\phi(x_n) + b) \geq 1, n = 1, 2, ..., N \tag{7}$$

The optimization goal is quadratic, the constraint is linear, and the whole is a convex quadratic programming problem.

Then we introduce Lagrangian multiplier $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]^T \geq 0$, to relax the objective function to be an objective function only subjected to a, we get Lagrange function:

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{n=1}^{N} \alpha_n (y_n (w^T \phi(x_n) + b) - 1) \tag{8}$$

Then the optimal solution would be:

$$p^* = \min_{w,b} \max_{\alpha \geq 0} L(w, b, \alpha) \tag{9}$$

The dual problem would be:

$$d^* = \max_{\alpha \geq 0} \min_{w,b} L(w, b, \alpha) \tag{10}$$

Original problem is difficult, thus we focus on dual problem first. We know that:

$$d^* \leq p^* \tag{11}$$

And when Slater's condition and KKT condition satisfy, $d^* = p^*$. We solve the dual problem.

First we fixed α, we optimize w and b.

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{n=1}^{N} \alpha_n y_n \phi(x_n) \quad = 0 \tag{12}$$

Then we can get

$$w = \sum_{n=1}^{N} \alpha_n y_n \phi(x_n) \tag{13}$$

And for b

$$\frac{\partial L(w, b, \alpha)}{\partial b} = - \sum_{n=1}^{N} \alpha_n y_n \quad = 0 \tag{14}$$

Then we can get

$$\sum_{n=1}^{N} \alpha_n y_n = 0 \tag{15}$$

Then we plug in w into $L(w, b, \alpha)$

$$
\begin{aligned}
L(w, b, \alpha) = & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi^T(x_i) y_i^T \alpha_i^T \alpha_j y_j \phi(x_j) \\
& - \sum_{i=1}^{N} \alpha_i (y_i (\sum_{j=1}^{N} \phi^T(x_j) y_j^T \alpha_j^T \phi(x_i) + b) - 1) \\
= & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi^T(x_i) y_i^T \alpha_i^T \alpha_j y_j \phi(x_j) \\
& - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i y_i \phi^T(x_j) y_j^T \alpha_j^T \phi(x_i) + \sum_{i=1}^{N} \alpha_i \\
= & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi^T(x_i) y_i^T \alpha_i^T \alpha_j y_j \phi(x_j)
\end{aligned} \tag{16}
$$

Then we can solve this problem by SMO(Sequential Minimal Optimization). When we get α^*, we can get

$$
w^* = \sum_{n=1}^{N} \alpha_n^* y_n \phi(x_n) \tag{17}
$$

$$
-1 = \frac{\max_{i, y_i=-1} w^{*T} x_i}{2} + b \tag{18}
$$

$$
1 = \frac{\min_{j, y_j=1} w^{*T} x_j}{2} + b \tag{19}
$$

$$
b^* = -\frac{\max_{i, y_i=-1} w^{*T} x_i + \min_{j, y_j=1} w^{*T} x_j}{2} \tag{20}
$$

1.2 Slater's condition

An optimization problem is like:

$$
\begin{aligned}
& \min f(x) \\
s.t. \quad & h_i(x) = 0, i = 1, 2, ..., p \\
& g_j(x) \leq 0, j = 1, 2, ..., q \\
& x \in X \subset R^n
\end{aligned} \tag{21}
$$

Where $f(x)$ is convex, and each $g_j(x)$ is convex, then if there exists an $x^* \in relint(D)$(the relative interior), such that

$$
\begin{aligned}
& h_i(x) = 0, i = 1, 2, ..., p \\
& g_j(x) < 0, j = 1, 2, ..., q
\end{aligned} \tag{22}
$$

Then saddle point exists, and strong duality holds. And when KKT condition satisfies, the saddle point is the optimal solution of original problem.

1.3 KKT condition

An optimization problem is like:

$$\begin{aligned} & \min f(x) \\ s.t. \quad & h_i(x) = 0, i = 1, 2, ..., p \\ & g_j(x) \leq 0, j = 1, 2, ..., q \\ & x \in X \subset R^n \end{aligned} \tag{23}$$

convex optimization: X is a convex set, $f : X \to R$ is a convex function, find a solution $x^* \in X$ that satisfies for every $x \in X$ have $f(x^*) \leq f(x)$.

$$\begin{aligned} & h_j(x^*) = 0, j = 1, 2, ..., p \\ & g_k(x^*) \leq 0, k = 1, 2, ..., q \\ & \nabla f(x^*) + \sum_{j=1}^{p} \lambda_j \nabla h_j(x^*) + \sum_{k=1}^{q} \mu_k \nabla g_k(x^*) = 0 \\ & \lambda_i \neq 0, \mu_k \geq 0, \mu_k g(x^*) = 0 \end{aligned} \tag{24}$$

When we consider SVM, we can find that the inequality condition holds.

$$g_n(w^*, b^*, \alpha^*) = 1 - y_n({w^*}^T \phi(x_n) + b^*) \leq 0, n = 1, 2, ..., N \tag{25}$$

And there is no equality condition $h(w^*, b^*, \alpha^*)$.

Then we know in SVM problem, α_n^* is the μ_k, we know that when we get optimal w^*, b^* and α^*:

$$\nabla f(w^*, b^*, \alpha^*) + \sum_{n=1}^{N} \mu_n \nabla g_n(w^*, b^*, \alpha^*) = \quad \nabla L(w^*, b^*, \alpha^*) = 0 \tag{26}$$

$$\alpha_n^* \geq 0, n = 1, 2, ..., N \tag{27}$$

Now we only need to prove that for optimal w^*, b^* and α^* that $\alpha_n g_n(w^*, b^*, \alpha^*) = 0, n = 1, 2, ..., N$ holds. We get optimal α^* for problem $\max_{\alpha \geq 0} L(w^*, b^*, \alpha)$ by letting $\nabla_\alpha L(w^*, b^*, \alpha) = 0$, then we can get that

$$\sum_{n=1}^{N} \alpha_n^* g_n(w^*, b^*, \alpha^*) = 0, n = 1, 2, ..., N \tag{28}$$

Since $\alpha_n^* \geq 0, n = 1, 2, ..., N$, and $g_n(w^*, b^*, \alpha^*) = 1 - y_n({w^*}^T \phi(x_n) + b^*) \leq 0, n = 1, 2, ..., N$, we can get

$$\alpha_n^* g_n(w^*, b^*, \alpha^*) = 0, n = 1, 2, ..., N \tag{29}$$

Then for the optimal solution w^*, b^* and α^* of SVM, the KKT condition holds.

1.4 Sequential Minimal Optimization for SVM

SMO is an algorithm to solve quadratic programming problem of SVM.

1. First find a pair α_i and α_j to update.
2. Fixed all the parameters except α_i and α_j, solve optimization problem to update α_i and α_j.

The optimization problem is

$$\begin{aligned} \max_{\alpha_i,\alpha_j} L(w^*, b^*, \alpha) &= \max_{\alpha_i,\alpha_j} \Big(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi^T(x_i) y_i^T \alpha_i^T \alpha_j y_j \phi(x_j) \Big) \\ s.t. \quad & \sum_{n=1}^{N} \alpha_n y_n = 0 \\ & \alpha_n \geq 0, n = 1, 2, ..., N \end{aligned} \tag{30}$$

When we fixed all the other parameters except α_i and α_j, the constraints would be

$$\begin{aligned} \alpha_i y_i + \alpha_j y_j &= c \\ \alpha_i \geq 0, \alpha_j &\geq 0 \\ c &= -\sum_{k \neq i,j} \alpha_k y_k \end{aligned} \tag{31}$$

The optimization problem would be:

$$\begin{aligned} \max_{\alpha_i,\alpha_j} L(w^*, b^*, \alpha) = \max_{\alpha_i,\alpha_j} \Big(& \alpha_i + \alpha_j \\ & - \frac{1}{2} \phi^T(x_i) y_i^T \alpha_i^T \sum_{k \neq i,j} \alpha_k y_k \phi(x_k) - \frac{1}{2} \phi^T(x_j) y_j^T \alpha_j^T \sum_{k \neq i,j} \alpha_k y_k \phi(x_k) \\ & - \frac{1}{2} \sum_{k \neq i,j} \phi(x_k)^T y_k^T \alpha_k^T \alpha_i y_i \phi(x_i) - \frac{1}{2} \sum_{k \neq i,j} \phi(x_k)^T y_k^T \alpha_k^T \alpha_j y_j \phi(x_j) \\ & - \frac{1}{2} \phi(x_i)^T y_i^T \alpha_i^T \alpha_j y_j \phi(x_j) - \frac{1}{2} \phi(x_j)^T y_j^T \alpha_j^T \alpha_i y_i \phi(x_i) \\ & - \frac{1}{2} \phi(x_i)^T y_i^T \alpha_i^T \alpha_i y_i \phi(x_i) - \frac{1}{2} \phi(x_j)^T y_j^T \alpha_j^T \alpha_j y_j \phi(x_j) \Big) \end{aligned} \tag{32}$$

$$s.t. \quad \alpha_i y_i + \alpha_j y_j = c \tag{33}$$

$$\alpha_i \geq 0, \alpha_j \geq 0 \tag{34}$$

$$c = -\sum_{k \neq i,j} \alpha_k y_k \tag{35}$$