Supporting Vector Machine

supporting vector

Dataset: $D = \{(x_1, y_1), \ldots, (x_m, y_m)\}, y_i \in \{-1, 1\}$.

To find a plane that separates these two classes of data points.

Define the hyperplane as: $w^T x + b = 0$,then the distance from point x to such plane is:

$$r = \frac{wx+b}{||w||} \tag{1}$$

suppose this plane can separate those two classes, then:

$$\begin{cases} w^T x_i + b \geq +1, y_i = 1 \\ w^T x_i + b \leq -1, y_i = -1 \end{cases} \tag{2}$$

Those points being the nearest ones satisfy the equal sign in (2), and they are called the **supporting vectors**. The sum of distance between two supporting vectors in different classes and the hyperplane is:

$$\gamma = \frac{2}{||w||} \tag{3}$$

γ is called the "**margin**".

> Now, the problem is find the max margin.

That is:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}||w||^2 \\ &s.t. \, y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, m \end{aligned} \tag{4}$$

And this is the **SVM model**.

Dual problem

To solve (4), we use Lagrange multiplier method as:

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(w^T x_i + b)) \tag{5}$$

with $\alpha = (\alpha_1, \ldots, \alpha_m)^T$,then:

$$\begin{aligned} w &= \sum_{i=1}^{m} \alpha_i y_i x_i \\ 0 &= \sum_{i=1}^{m} \alpha_i y_i \end{aligned} \tag{6}$$

By substituting (6) to (5), we can obtain the dual problem of (4):

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j$$
$$s.t. \sum_{i=1}^{m} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \ldots, m \tag{7}$$

$x_i^T x_j$ is the inner product of the vector x_i and x_j, it can also be written as $< x_i, x_j >$.

By solving the above problem, we can obtain our SVM model as:

$$f(x) = w^T x + b = \sum_{i=1}^{m} \alpha_i y_i x_i x + b \tag{8}$$

KKT condition

Due to the inequality constraints in (4), it should meet the KKT condition:

$$\begin{cases} \alpha_i \geq 0; \\ y_i f(x_i) - 1 \geq 0; \\ \alpha_i (y_i f(x_i) - 1) = 0 \end{cases} \tag{9}$$

From (9), we know that for α_i and $y_i f(x_i) - 1$, at least one of them is zero. If α_i is zero, it makes no contribution in (8), so there must has $y_i f(x_i) = 1$, corresponding to those **supporting vectors**.

> Thus, only Supporting Vectors matters to the final model.

SMO (Sequential Minimal Optimization) algorithm

To solve (7), SMO is a good algorithm.

> - choose a pair of (α_i, α_j) to be updated.
> - Fix other α, to solve (7). ($\sum_{i=1}^{m} \alpha_i y_i = 0$)