Support Vector Machine

$$h(x) = \mathrm{sign}(\underline{w^T x + b})$$

- Linear discriminant function
- represents a straight line in 2D
- represents a plane in 3D
- represents a hyperplane in >3D

w: a vector perpendicular to the hyperplane

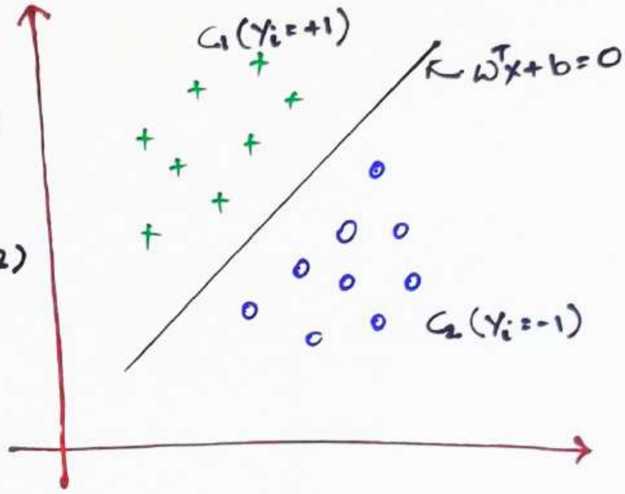
b: position of the hyperplane in the d-dimensional space

$h(x) = w^T x + b = 0$

$h(x_i) = w^T x_i + b > 0$ (x_i is in positive side of the hyperplane) —(1)

$= w^T x_i + b < 0$ (x_i is in negative side of the hyperplane) —(2)

$= w^T x_i + b = 0$ (x_i is on the hyperplane)



Dataset:

$\{x_i, y_i\}$ $i = 1 \ldots m$

where, $x \in \mathbb{R}^d$

$y \in \{+1, -1\}$

Combining Eq. 1 and Eq. 2

$$y_i(w^T x_i + b) > 0$$

We want each data point to be at least γ distance away from the hyperplane.

hyperplane: $w^T x + b = 0$

distance of x from the hyperplane

$$\frac{w^T x + b}{\|w\|}, \text{ so,}$$

$$\frac{w^T x + b}{\|w\|} \geq \gamma$$

w represents the scaling of the hyperplane, scaling doesn't affect learning of w. So,

$$w^T x + b \geq \gamma \|w\| \quad \text{by scaling, assume } \gamma\|w\| = 1$$

$$w^T x_i + b \geq 1 \quad \text{if } x_i \in C_1$$

$$w^T x_i + b \leq -1 \quad \text{if } x_i \in C_2$$

So,

$$y_i(w^T x_i + b) \geq 1$$

To build SVM, we want to maximize distance between the linear hyperplane and input data points. So, we have to

$$\text{maximize } \frac{2}{\|w\|}$$

$$\text{minimize } \|w\|$$

Optimization Problem:

$$\boxed{\begin{array}{l} \text{minimize } \frac{1}{2}\|w\|^2 \\ \text{subject to } y_i(w^T x_i + b) \geq 1 \quad \text{for } i = 1 \ldots m \end{array}} \quad \text{——— (3)}$$

Solution using Lagrange multipliers:

The above problem is a constrained optimization problem. So, this can be solved using Lagrange multiplier and KKT condition. (Karush-Kuhn-Tucker)

$$\mathcal{L}(w,b) = \frac{1}{2} w^T w - \sum \alpha_i [y_i(w^T x_i + b) - 1] \qquad \alpha_i : \text{Lagrange multiplier } (\alpha_i \geq 0)$$

$$= \frac{1}{2} w^T w - \sum \alpha_i y_i w^T x_i - \sum \alpha_i y_i b + \sum \alpha_i \quad \text{——— (4)}$$

$$\frac{\partial \mathcal{L}(w,b)}{\partial w} = w - \sum \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum \alpha_i y_i x_i} \quad \text{——— (5)}$$

$$\frac{\partial \mathcal{L}(w,b)}{\partial b} = -\sum \alpha_i y_i = 0$$

$$\boxed{\sum \alpha_i y_i = 0} \quad \text{—— (6)}$$

From Eq. 4.

$$\mathcal{L}(w,b) = \frac{1}{2} w^T w - \sum \alpha_i y_i w^T x_i - \underbrace{\sum \alpha_i y_i b}_{=0} + \sum \alpha_i$$

using Eq. 5 & Eq. 6.

$$= \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j (x_i^T x_j) - \sum \alpha_i \alpha_j y_i y_j (x_i^T x_j) + \sum \alpha_i$$

$$\boxed{= \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j (x_i^T x_j)} \quad \text{—— (7)}$$

— DUAL

Putting everything together,

$$\alpha_i \geq 0 \qquad \sum \alpha_i y_i = 0 \qquad w = \sum \alpha_i y_i x_i$$

Using KKT condition,

$$\alpha_i [y_i (w^T x_i + b) - 1] = 0$$

If $\alpha_i > 0$ $\quad y_i (w^T x_i + b) = 1$ These x_i's are called <u>Support Vector</u>.
There exist at least 1 support vector on both sides of hyperplane

If $\alpha_i = 0$ $\quad y_i (w^T x_i + b) > 1$

SVM Classifier:

$$\boxed{h(x_i) = \text{sign}\left(\sum \alpha_j y_j x_j^T x_i + b\right)}$$