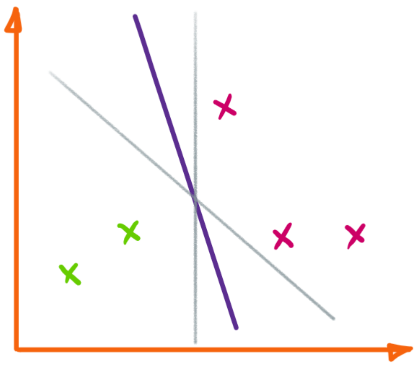
Max Margin Classifier for SVM

Zengchang Qin (Ph.D)

Max Margin

Given a binary classification problem, we need to find an optimal hyperplane.



There exists more than one hyperplane to satisfy the condition of correct classification, which one is the best?

1) Using the linear model $h_w(x) = w^T x + b$

we hope to map data $x^{(i)}$: $h_w(x^{(i)}) \rightarrow y^{(i)}$

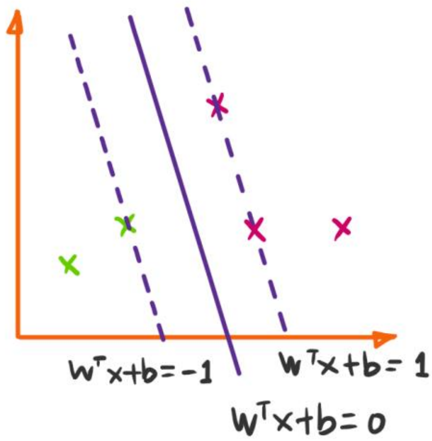
where $y^{(i)} \in \{-1, +1\}$. given training data

$$D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}) \cdots (x^{(N)}, y^{(N)})\}$$

By correctly classify all the data, we have:

$$y^{(i)} h_w(x^{(i)}) > 0$$

By rescaling the data, we can always find two hyperplanes, the space between them is the margin of two classes.



The data on the margin satisfy:

$$w^Tx+b=1 \quad \text{or} \quad w^Tx+b=-1$$

Therefore, if $x^{(i)}$ is correctly classified, then

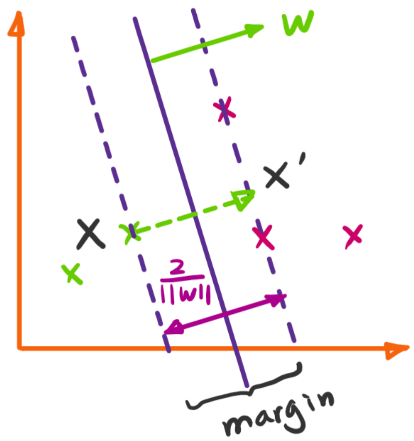
$$y^{(i)}(w^Tx^{(i)}+b) \geq 1$$

we need to find the hyperplane with maximum margin with constraints of classifying data correctly.

2) If x is on one margin such that

$$w^Tx+b=-1$$

we can map x to the other margin in the direction of w, with x' satisfying: $w^Tx'+b=1$



The margin is $\lambda\|w\| = \|x' - x\|$

Since $w^T x' + b = 1$, $x' = x + \lambda w$

we yield $w^T(x + \lambda w) + b = 1$

$$\Rightarrow w^T x + \lambda w^T w + b = 1$$

$$\Rightarrow w^T x + b = 1 - \lambda \|w\|^2$$

Because $w^T x + b = -1$

we then have : $\lambda \|w\|^2 = 2$

$$\Rightarrow \lambda = 2 / \|w\|^2$$

By substituting λ, the margin is :

$$\lambda \|w\| = 2/\|w\|^2 \cdot \|w\| = \frac{2}{\|w\|}$$

3) The classification problem becomes an optimization problem:

$$\begin{cases} \max_{w} \frac{2}{\|w\|} \\ \text{s.t. } y^{(i)} w^T x^{(i)} \geq 1 \quad \text{for } i = 1 \cdots N \end{cases}$$

For mathematical convenience, we can rewrite it as a minimization problem:

$$\begin{cases} \min_{w} \|w\|^2/2 \\ \text{s.t.: } y^{(i)} w^T x^{(i)} + b \geq 1, \quad \text{for } i = 1, 2 \cdots N \end{cases}$$

Dual Form

5) Introducing Lagrange multiplier, the constraint becomes

$$\max_{\alpha_i} \alpha_i [1 - y_i (w x_i + b)] \quad , \alpha_i \geq 0$$

If x_i is correctly classified, $\alpha_i = 0$ in order to maximize the term, otherwise (misclassified) $\alpha_i \to \infty$ in order to penalize the new constructed Lagrangian function:

$$\min_{w,b} L = \frac{1}{2}\|w\|^2 + \max_{\alpha_i} \sum_i \alpha_i [1 - y_i (w x_i + b)] \quad (6)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C$$

$$\min L = \max_{\alpha_i} \min_{w,b} \frac{1}{2}\|w\|^2 + \sum_i \alpha_i [1 - y_i (w x_i + b)] \quad (7)$$

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_i \alpha_i y_i x_i & (8) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_i \alpha_i y_i = 0 & (9) \end{cases}$$

6) Dual Form of Lagrange function L

$$L(w,b) = \frac{1}{2}\|w\|^2 + \Sigma_i \alpha_i [1 - y_i(wx_i + b)]$$

Substitute Eq. (8)

$$L = \frac{1}{2}(\Sigma_i \alpha_i y_i x_i)(\Sigma_j \alpha_j y_j x_j) + \Sigma_i \alpha_i [1 - y_i(\Sigma_j \alpha_j y_j x_j x_i + b)]$$

$$= \underline{\frac{1}{2}(\Sigma_i \alpha_i y_i x_i)(\Sigma_j \alpha_j y_j x_j)} + \Sigma_i \alpha_i \underline{- \Sigma_i \alpha_i y_i \Sigma_j \alpha_j y_j x_j x_i} - \underbrace{\Sigma_i \alpha_i y_i b}_{\Sigma \alpha_i y_i = 0}$$

$$= \Sigma_i \alpha_i - \frac{1}{2}\Sigma_i \Sigma_j \alpha_i \alpha_j y_i y_j x_i x_j$$

The optimization of $L(\alpha)$ can be solved by quandritic programing.