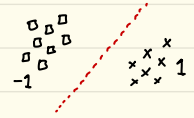


Support Vector Machines (SVMs)

among the best off-the-shelf supervised learning algorithms

They rely on the notion of 'margin' and try to find a classifier with a large 'margin'

SVM is closely related to logistic regression in that both try to find a hyperplane that separates the data



The differences, however, are: (1) The cost functions that are optimized in SVM and LR are different:

SVM : maximizes the worst 'margin'

LR : maximizes the likelihood

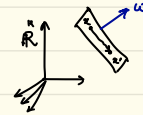
(2) SVM is more versatile and can handle data that are not linearly separable, instead can be separated by a nonlinear decision boundary.



* Given a dataset $D = \{(x_i^T, y_i), \dots, (x_n^T, y_n)\}$ where $x_i^T \in \mathbb{R}^k$ and $y_i \in \{-1, +1\}$, find a decision boundary $\tilde{w}x + b = 0$ that best separates the data with 'maximum worst margin'.

For now, let's assume only two classes, and data separable by a hyperplane.

(I) $\underset{\mathbb{R}^k}{\tilde{w}}^T \underset{\mathbb{R}^k}{x} + \underset{\mathbb{R}}{b} = 0 \rightarrow$ specifies the equation of a hyperplane:

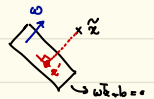


Assume x, x' are on the hyperplane $\begin{cases} \tilde{w}x + b = 0 \\ \tilde{w}x' + b = 0 \end{cases} \Rightarrow \tilde{w}^T(x - x') = 0 \rightarrow$ Thus, \tilde{w} is the normal to the hyperplane and b is the offset from origin.

(II) What is the distance of a point \tilde{x} to hyperplane $\tilde{w}x + b = 0$.

$\begin{cases} \tilde{x} - x' \text{ is orthogonal to the plane} \rightarrow \tilde{x} - x' \text{ is parallel to } \tilde{w} \rightarrow \tilde{x} - x' = \alpha \tilde{w} \\ x' \text{ is on the plane} \rightarrow \tilde{w}^T x' + b = 0 \end{cases} \quad (2)$

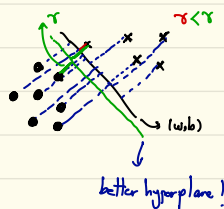
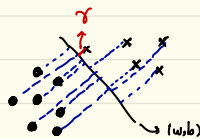
$$\downarrow \\ x' = \tilde{x} - \alpha \tilde{w} \quad (1)$$



$$(1), (2) \Rightarrow \tilde{w}^T(\tilde{x} - \alpha \tilde{w}) + b = 0 \rightarrow \tilde{w}^T \tilde{x} + b = \alpha \|\tilde{w}\|^2 \rightarrow \alpha = \frac{\tilde{w}^T \tilde{x} + b}{\|\tilde{w}\|^2}$$

The distance of \tilde{x} to $w^T x + b = 0$ is $\|\tilde{x} - x'\|_2 = \|\alpha w\|_2 = |\alpha| \|w\|_2 = \frac{|w^T \tilde{x} + b|}{\|w\|_2}$

* Given $D = \{(x^1, y^1), \dots, (x^N, y^N)\}$, we have



Distance of each x_i from $w^T x + b = 0$ is $\frac{|w^T x_i + b|}{\|w\|_2}$

$$\text{Let } \gamma(w, b) \triangleq \min_{i=1, \dots, N} \frac{|w^T x_i + b|}{\|w\|_2} \Rightarrow \forall i=1, \dots, N : \frac{y_i (w^T x_i + b)}{\|w\|_2} \geq \gamma$$

We want to find a hyperplane (w, b) for which $\gamma(w, b)$ (the minimum margin is maximized)

$$\begin{cases} \max_{w, b, \gamma} \gamma \\ \text{s.t. } \frac{y_i (w^T x_i + b)}{\|w\|_2} \geq \gamma, \forall i=1, \dots, N \end{cases}$$

$$\text{Let } \bar{\gamma} \triangleq \|w\|_2 \gamma$$

We can write the above maximization in the following equivalent form:

$$\begin{cases} \max_{w, b, \bar{\gamma}} \frac{\bar{\gamma}}{\|w\|_2} \\ \text{s.t. } y_i (w^T x_i + b) \geq \bar{\gamma}, \forall i=1, \dots, N \end{cases}$$

If $(w^*, b^*, \bar{\gamma}^*)$ a solution, then $(\alpha \bar{\gamma}^*, \alpha w^*, \alpha b^*)$ also a solution for any $\alpha > 0$.

We can fix this by enforcing $\bar{\gamma}^* = 1 \rightarrow (w^*, b^*, 1) \rightarrow$ can't choose any $\alpha \neq 1$.

$$\rightarrow \begin{cases} \max_{w, b} \frac{1}{\|w\|_2} \\ \text{s.t. } y_i (w^T x_i + b) \geq 1, \forall i=1, \dots, N \end{cases} \rightarrow \begin{cases} \min_{w, b} \frac{1}{2} \|w\|_2^2 \\ \text{s.t. } y_i (w^T x_i + b) \geq 1, \forall i=1, \dots, N \end{cases} \quad \leftarrow \text{Vanilla SVM}$$