Support Vector Machines

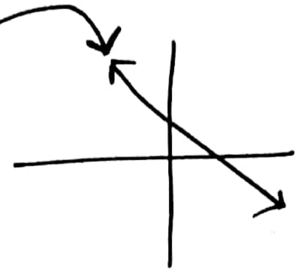
$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix}$; $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix}$

w_0 : bias, $x_0 = +1$

- Eq^n of line: $\bar{w} \cdot \bar{x} + x_0 = 0$

  $\Rightarrow w^T x + x_0 = 0$

  Normal Vector is $\bar{w}$.



- Distance bet^n 2 || lines.

  $ax + by + c_1 = 0$; $ax + by + c_2 = 0$

  $$d = \frac{|c_2 - c_1|}{\sqrt{a^2 + b^2}}$$

Decision Surface

$y = w^T x + w_0$

For Surface;

$y = 0$

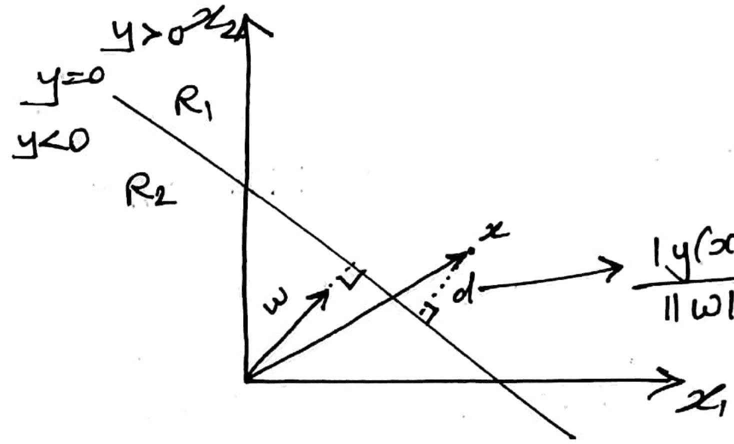
$y > 0$; +ve sample

$y < 0$; −ve sample

- Distance bet^n point & line

  (x_0, y_0) $ax + by + c = 0$

  $$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$



$$\frac{|y(x)|}{\|w\|}$$

$\text{target}_n \cdot y(x_n) > 0$

Distance : $\dfrac{|y(x_n)|}{\|w\|} = \dfrac{t_n\, y(x_n)}{\|w\|} = \dfrac{t_n(w^T \phi(x) + w_0)}{\|w\|}$

Distance from Origin $(0,0)$ to decision surface $= \dfrac{|w_0|}{\|w\|}$

MIT

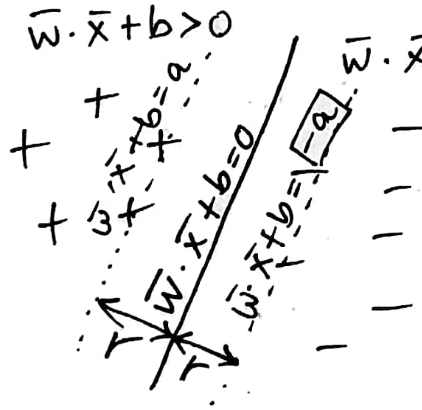
$\bar{w} \cdot \bar{u} \geq c$

$\bar{w} \cdot \bar{u} + b \geq 0$

Decision Rule

+ve

$\bar{w} \cdot \bar{x} + b > 0$ $\qquad$ $\bar{w} \cdot \bar{x} + b < 0$



confidence

$(\bar{w} \cdot \bar{x}_j + b) y_j$

Margin:

$$r = \frac{a}{\|w\|}$$

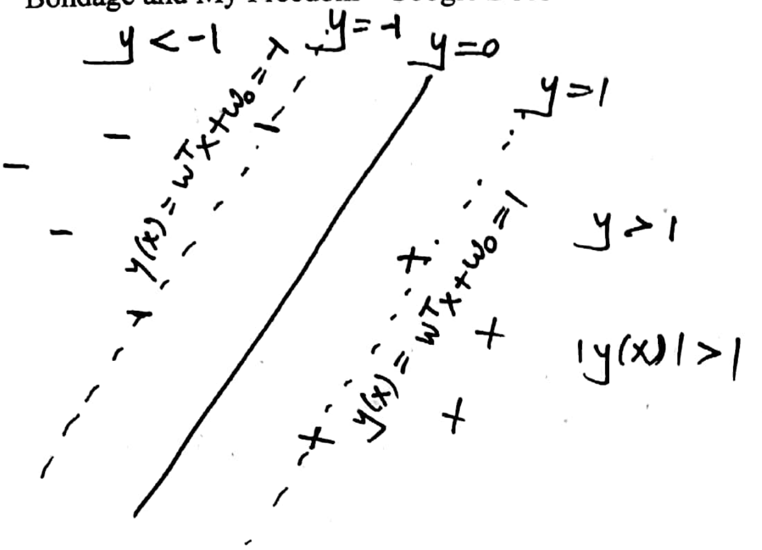
Dist bet^n || lines

$$2r = \frac{|2a|}{\|w\|}$$

$$\Rightarrow r = \frac{|a|}{\|w\|}$$

set $a = 1$

Mathematically conuinient.



Better Confidence

target $* \, y(x)$

Distance on either side $= \frac{1}{||w||}$

MAXIMIZE: $\frac{1}{||w||}$

Constraints

target x predicted ≥ 1

$t_n(w^T x_n + b) \geq 1$

Maximize: $\frac{1}{||w||} \longrightarrow$ Minimize: $||w|| \longrightarrow$ Minimize: $\frac{1}{2}||w||^2$

constraints: $t_n(w^T \phi(x_n) + b) \geq 1$

To find optima, given constraints $\Rightarrow$ Lagrange Multiplier.

Lagrange: $\alpha = 0$: constraint ineffective, $\alpha > 0$ effective

Primal: minimize$_{w,b}$ $\frac{1}{2} w \cdot w$

s.t. $y_j(w \cdot x_j + b) \geq 1$

$y_j(w \cdot x_j + b) - 1 \geq 0$

W: wt on features

Lagrange

$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum \alpha_j [(w \cdot x_j + b) y_j - 1]$

α: wt on training points

$\alpha_j \geq 0, \forall j$

minimize (x^2)

constraint $x \geq b$

$x - b \geq 0$

$L = x^2 - \alpha(x - b)$

$\alpha \geq 0$

differentiate L.

$\max_\alpha d(\alpha) \Rightarrow \min_x L(x, \alpha)$

such that $\alpha \geq 0$

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{n=1}^{N} a_n\left[t_n(w^T\phi(x_n)+b)-1\right]$$

$a = (a_1, a_2, \dots a_n)^T$

partial derivative wrt w.

$$w - \sum a_n t_n \phi(x_n) = 0 \Rightarrow w = \sum_{n=1}^{N} a_n t_n \phi(x_n)$$

p.d wrt b

$$0 - \sum a_n t_n = 0 \Rightarrow \sum_{n=1}^{N} a_n t_n = 0$$

Famous Kernels

Poly: $7q_1^2 q_2^3 q_4$

$K(q,q') = (1+q\cdot q')^K$

Sigmoid: $\tanh(q_1 + 3q_2)$

$K(q,q') = \tanh(aq\cdot q' + b)$

gaussian $e^{-\frac{1}{2}(q_1-5)^2}$

$$K(q,q') = e^{\frac{-\|q-q'\|^2}{\sigma^2}}$$

Putting these values back;

$$L = \frac{1}{2}\left(\sum_{n=1}^{N} a_n t_n \phi(x_n)\right)\left(\sum_{m=1}^{N} a_m t_m \phi(x_m)\right) - \sum_{n=1}^{N} a_n t_n \phi(x_n)\left(\sum_{m=1}^{N} a_m t_m \phi(x_m)\right) - b\sum a_n t_n + \sum a_n$$

$$= \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} a_n a_m t_n t_m \underbrace{\phi(x_n)\cdot\phi(x_m)}_{\phi(x_n)^T\phi(x_m)}$$

Original Eq^n

$y(x) = w^T x + \cancel{w_0}\, b \geq 0$

$y(x) = \sum a_n t_n \phi(x_n)\phi(x) + b$

KKT conditions

$a_n \geq 0$

$t_n y(x_n) - 1 \geq 0$

$a_n\{t_n y(x_n) - 1\} = 0$

Either

$a_n = 0$

OR

$t_n y(x_n) = 1$

Support Vectors

Kernel Trick

$K(x, x') = \phi(x)^T \phi(x) = [redacted] = \phi(\bar{x}) \cdot \phi(\bar{x'})$

We only need, dot product in space

$X = \mathbb{R}^d: \phi: X \to Z$, poly Q order

$K(x, x') = (1 + x^T x')^Q$

$= (1 + x_1 x_1' + x_2 x_2' + \ldots x_d x_d')^Q$

$d = 10, Q = 100$

scale: $K(x, x') = (a x^T x' + b)^Q$

Given 2 points $x \,\&\, x' \in X$, we need $z^T z'$

$z^T z' = K(x, x')$ (kernel)

inner product of $x \,\&\, x'$

$x = (x_1, x_2)$ ϕ: 2^{nd} order

$z = \phi(x) = (1, x_1, x_2, x_1^2, x_2^2, x_1 x_2)$

$K(x, x') = z^T z'$

$$= \begin{bmatrix} 1 \\ x_1' \\ x_2' \\ x_1'^2 \\ x_2'^2 \\ x_1' x_2' \end{bmatrix} [1 \; x_1 \; x_2 \; x_1^2 \; x_2^2 \; x_1 x_2]$$

$$= 1 + x_1 x_1' + x_2 x_2' + x_1^2 x_1'^2 + x_2^2 x_2'^2 + x_1 x_2 x_1' x_2'$$

Consider $K(x, x') = (1 + x^T x')^2$ Only for support vectors

$$= \left(1 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [x_1' \; x_2']\right)^2$$

$$= (1 + x_1 x_1' + x_2 x_2')^2$$

$$= 1 + x_1^2 x_1^2 + x_2^2 x_2'^2 + 2 x_1 x_1' + 2 x_1 x_1' x_2 x_2' + 2 x_2 x_2'$$

$$\begin{pmatrix} 1 \\ x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 \\ \sqrt{2} x_2 \\ \sqrt{2} x_1 x_2 \end{pmatrix} (1 \; x_1'^2 \; x_2'^2 \; \sqrt{2} x_1' \; \sqrt{2} x_2' \; \sqrt{2} x_1' x_2')$$

Inner Product.

Slack Variables:

$\varepsilon_n \geq 0$ | dist to be moved

Correct boundary $\varepsilon_n = 0$.

otherwise

$\varepsilon_n = |t_n - y(x_n)|$

If on decision boundary: $1 \to \varepsilon_n$

$t_n y(x_n) \geq 1 - \varepsilon_n$

New Optimization

$$C \sum_{n=1}^{n} \varepsilon_n + \frac{1}{2} \|w\|^2$$

tradeoff slack penalty & margin condition

$$L = \frac{1}{2}\|w\|^2 + C \sum \varepsilon_n - \sum a_n (t_n y(x_n) - 1 + \varepsilon_n) - \sum \mu_n \varepsilon_n$$

$a_n \geq 0$; $t_n(y(x_n)) - 1 + \varepsilon_n \geq 0$, $\mu_n \geq 0$, $\varepsilon_n \geq 0$,

$a_n(t_n y(x_n) - 1 + \varepsilon_n) = 0$ $\mu_n \varepsilon_n = 0$

$w = \sum a_n t_n \phi(x_n)$; $\sum a_n t_n = 0$; $a_n = C - \mu_n$

$\boxed{\text{pd} \quad w \quad b \quad \varepsilon_n}$

$$L = \sum a_n - \frac{1}{2} \sum \sum a_n a_m t_n t_m K(x_n, x_m)$$

$0 \leq a_n \leq C$; $\sum a_n t_n = 0$

margin support vector $0 < a_n < C$

$t_n(w^T x_n + b) = 1$

slack $= 0$

non margin

$t_n(w^T x_n + b) < 1$

wrong side

slack > 0

(4)