Kernel Machines

different models — inductive bias
- ⇒ different assumptions
- ⇒ different objective functions

SUPPORT VECTOR MACHINES (SVM)

↳ They do not care about probabilities or densities.

↳ Weights can be written in terms of training data points. — representer theorem

$$g(x) = \underbrace{w^T}_{1 \times D} \cdot \underbrace{x}_{D \times 1} + \underbrace{w_0}_{1 \times 1}$$

$\theta = \{w, w_0\}$

\# of parameters $= D+1$

$$w = \sum_{i=1}^{N} \alpha_i x_i$$

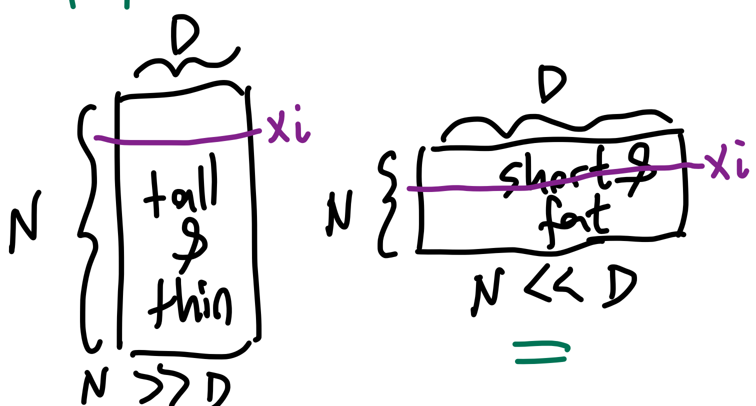
α_i's are mostly zero.

$$g(x) = w^T \cdot x + w_0$$

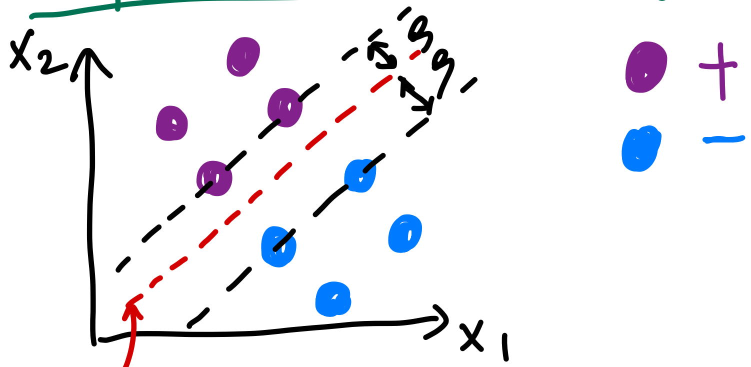
$$= \Big[\underbrace{\sum_{i=1}^{N} \alpha_i x_i}_{\text{training data points}}\Big]^T \cdot \underbrace{x}_{\text{test point}} + w_0 = \sum_{i=1}^{N} \alpha_i [x_i^T \cdot x] + w_0$$

$\theta = \{\alpha_1, \alpha_2, \ldots, \alpha_N, w_0\}$

\# of parameters $= N+1$



Optimal Separating Hyperplane



$x_1 + x_2 \geq -5$

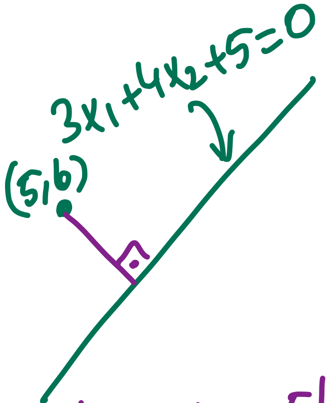
$-x_1 - x_2 \leq +5$

$3x_1 + 4x_2 - 5 = 0$
$-3x_1 - 4x_2 + 5 = 0$
$6x_1 + 8x_2 - 10 = 0$
equivalent

$w^T.x + w_0 = 0$ $\quad \theta = \{w, w_0\}$

0 1

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \quad x_i \in \mathbb{R}^D \quad y_i \in \{-1, +1\}$$



$$\boxed{\begin{aligned} (w^T.x_i + w_0)\, y_i &\geq +1\,(y_i) \quad \text{if } y_i = +1 \\ (w^T.x_i + w_0)\, y_i &\leq -1\,(y_i) \quad \text{if } y_i = -1 \end{aligned}}$$

$$\Downarrow$$

$$y_i(w^T.x_i + w_0) \geq 1 \quad \forall i$$

$$\frac{|3x_1+4x_2+5|}{\sqrt{3^2+4^2}} = \frac{44}{5} = 8.8$$

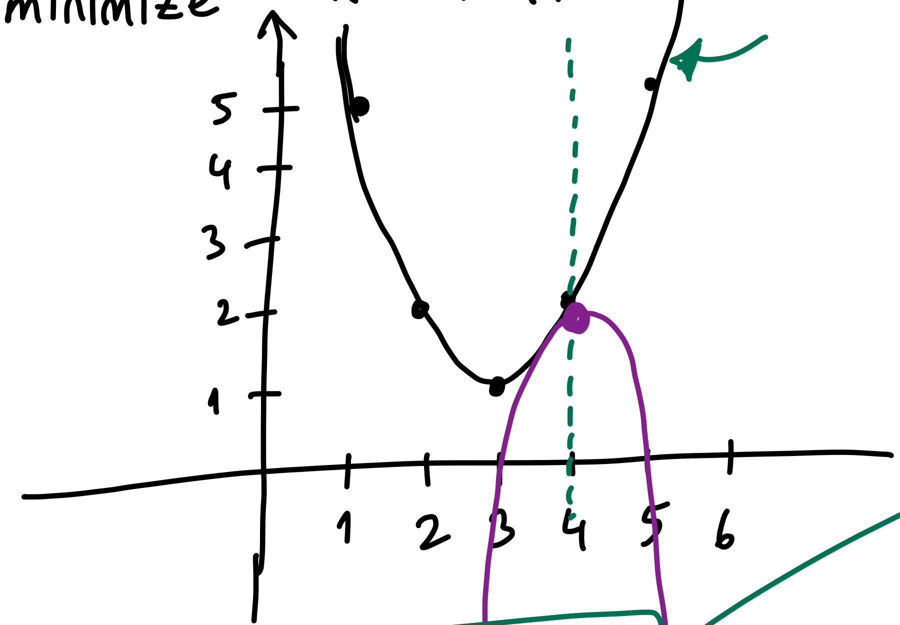
$$\frac{|w^T.x_i + w_0|}{\|w\|_2} = \frac{y_i(w^T.x_i + w_0)}{\|w\|_2} \geq \rho \Rightarrow y_i(w^T.x_i + w_0) \geq \rho \|w\|_2$$

to obtain a unique solution $\Rightarrow \rho \|w\|_2 = 1$

minimize $x^2-6x+10$

$$\frac{\partial(x^2-6x+10)}{\partial x} = 2x-6 \Rightarrow \underline{x^* = 3}$$

$$\frac{\partial^2(x^2-6x+10)}{\partial x^2} = 2 \geq 0$$



maximize $-x^2+8x-14$

minimize $x^2-6x+10$
subject to: $x \geq 4$

$\Rightarrow$ minimize $x^2-6x+10$
subject to: $x-4 \geq 0$] λ

$$\text{Lagrangian}(x,\lambda) = x^2-6x+10-\lambda[x-4]$$ ↲ minimize this function

$$\frac{\partial[x^2-6x+10-\lambda x+4\lambda]}{\partial x} = 2x-6-\lambda = 0$$

$$\lambda = 2x-6$$

$$x = (\lambda+6)/2$$

$$x^2-6x+10-\overset{\lambda}{(2x-6)}(x-4) = x^2-6x+10-2x^2+14x-24$$

$$= -x^2+8x-14$$ ↲ maximize this function

$$-2x+8=0 \Rightarrow \underline{\underline{x^*=4}}$$

$\left[\, \|w\|_2\, g = 1 \,\right]$ maximize g $\Longleftrightarrow$ minimize $\|w\|_2$

$$\Downarrow$$

$$\left(\sqrt{w_1^2 + w_2^2 + \dots + w_D^2}\right)$$

minimize $\|w\|_2$ $\Longleftrightarrow$ minimize $\frac{1}{2}\|w\|_2^2$

$w^T = [w_1\ w_2 \dots w_D]$, $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix}$

$$\boxed{\begin{aligned} &\text{minimize } \tfrac{1}{2}\|w\|_2^2 \\ &\text{subject to: } y_i(w^T.x_i + w_0) \geq 1 \quad \forall i \end{aligned}}$$

$\alpha_i \rightarrow$ Lagrangian coefficient

$\rightarrow$ separation constraints

\# of constraints $= N$

\# of decision variables $= D+1$

$$\left\{ L_p = \frac{1}{2} w^T.w - \sum_{i=1}^{N} \alpha_i \left[\underbrace{y_i(w^T.x_i + w_0) - 1}_{\text{seperation const.}} \right] \right.$$

representer theorem

$$\frac{\partial L_p}{\partial w} = \frac{1}{\cancel{2}}.\cancel{2}.w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \Rightarrow \boxed{w = \sum_{i=1}^{N} (\alpha_i y_i)\, x_i}$$

$$\frac{\partial L_p}{\partial w_0} = -\sum_{i=1}^{N} \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{N} (\alpha_i y_i) = 0$$

$$\left(\sum_{i=1}^{N} a_i\right)\left(\sum_{i=1}^{N} a_i\right) \neq \sum_{i=1}^{N} a_i^2 \Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j$$

$$(a_1 + a_2)(a_1 + a_2) = a_1^2 + \underline{2a_1a_2} + a_2^2$$

$$\|w\|_2^2 = w^T.w = \left[\sum_{i=1}^{N}\alpha_i y_i x_i\right]^T\left[\sum_{j=1}^{N}\alpha_j y_j x_j\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T.x_j$$

$$L_p = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T.x_j - \sum_{i=1}^{N}\alpha_i\left[y_i\left[\left(\sum_{j=1}^{N}\alpha_j y_j x_j\right)^T.x_i + w_0\right] - 1\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T.x_j - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \cancel{\left(\sum_{i=1}^{N}\alpha_i y_i\right)w_0} + \sum_{i=1}^{N}\alpha_i$$

$$= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T.x_j$$

$$x_j^T.x_i = x_i^T.x_j$$

$$\boxed{\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{N}\overset{?}{\alpha_i} - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overset{?}{\alpha_i}\overset{?}{\alpha_j}\overset{\checkmark}{y_i}\overset{\checkmark}{y_j}\overset{\checkmark}{x_i^T}.\overset{\checkmark}{x_j} \\ &\text{subject to:} \quad \sum_{i=1}^{N}\overset{?}{\alpha_i}\overset{\checkmark}{y_i} = 0 \\ &\qquad\qquad\qquad \overset{?}{\alpha_i} \geq 0 \quad \forall i \end{aligned}}$$

\# of constraints = 1

\# of decision variables = N