Kernel Machines

different models [⇒ different assumptions
inductive bias [⇒ different objective functions

SUPPORT VECTOR MACHINES (SVM)

↳ They do not care about probabilities or densities.
↳ Weights can be written in $\underbrace{\text{terms of training data points.}}_{\text{representer theorem}}$

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

w: $D\times 1$, x: $D\times 1$, w_0: 1×1

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

\# of parameters $= D+1$

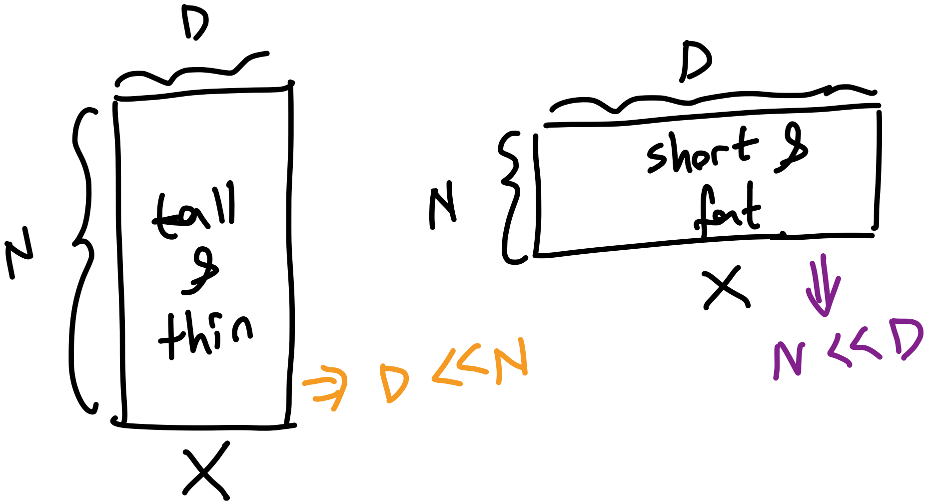
$$\boxed{w = \sum_{i=1}^{N} \alpha_i x_i}$$

↳ α_i's are mostly zero.

$$g(x) = w^T \cdot x + w_0$$
$$= \left(\sum_{i=1}^{N} \alpha_i \cdot x_i\right)^T \cdot x + w_0$$
$$= \sum_{i=1}^{N} \alpha_i \left[x_i^T \cdot x\right] + w_0$$

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

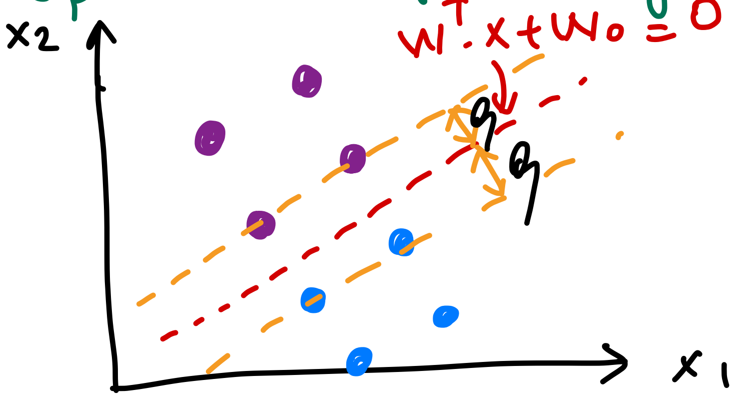
\# of parameters $= N+1$



Optimal Separating Hyperplane

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

$\begin{matrix} 1 & 2 \\ 0 & 1 \end{matrix}$



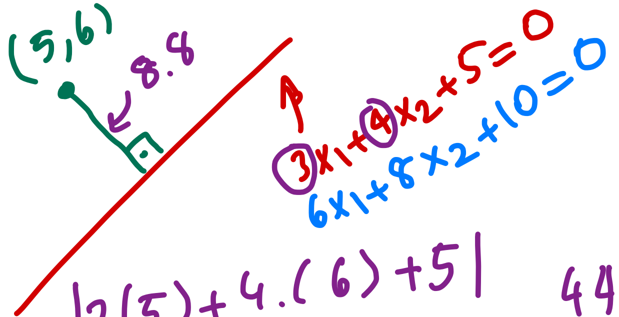
● +

● −

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

$-1\,(x_1 + x_2) \geq (5)\,-1$

$-x_1 - x_2 \leq -5$



$$\frac{|3(5) + 4.(6) + 5|}{\sqrt{3^2 + 4^2}} = \frac{44}{5} = 8.8$$

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

$$\Downarrow$$

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

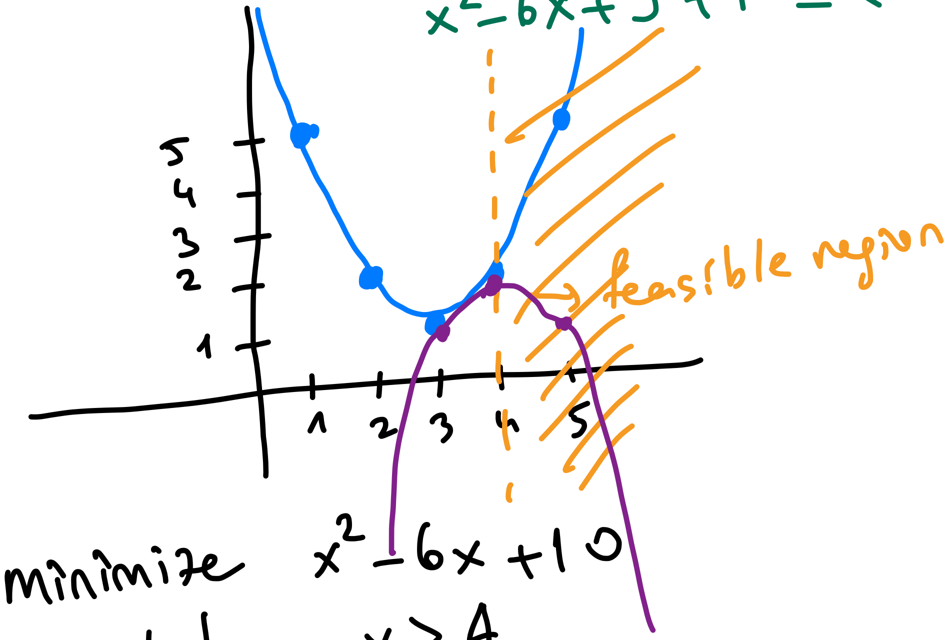
$$\frac{|w^T.x_i + w_0|}{\|w\|_2} = \frac{y_i(w^T.x_i + w_0)}{\|w\|_2} \geq \rho \quad \forall i$$

$$y_i(w^T.x_i + w_0) \geq \rho \|w\|_2 \quad \Leftarrow \text{separation constraint}$$

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

minimize $x^2 - 6x + 10$

$x^2 - 6x + 9 + 1 = (x-3)^2 + 1$



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

$x^* = 3$

$$\frac{\partial (2x-6)}{\partial x} = 2 > 0$$

→ strictly convex

minimize $x^2 - 6x + 10$

subject to: $x \geq 4$

minimize $x^2 - 6x + 10$

subject to $x - 4 \geq 0 \;] \; \lambda$

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

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

$[\lambda^* = 2x - 6]$ or $x^* = \left[\frac{\lambda + 6}{2}\right]$

$x^2 - 6x + 10 - (2x-6)(x-4) = x^2 - 6x + 10 - 2x^2 + 14x - 24$

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

$= -(x-4)^2 + 2$

$$\frac{\partial (-x^2 + 8x - 14)}{\partial x} = -2x + 8$$

$\underline{\underline{x^* = 4}}$

maximize δ = minimize $\|w\|_2$ [since $\delta \|w\|_2 = 1$]

$\sqrt{w_1^2 + w_2^2 + \dots + w_D^2}$

$\Downarrow$

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

$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^N$

PRIMAL

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

Lagrangian coefficient

α_i

$w_1^2 + \dots + w_D^2 = w^T.w$

# of constraints = N # of decision variables = D+1

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

$$\frac{\partial L_p}{\partial w} = \frac{1}{2}.2.\underset{D\times 1}{w} - \sum_{i=1}^N \underset{1\times 1}{\alpha_i} \underset{1\times 1}{y_i} . \underset{D\times 1}{x_i} = 0 \Rightarrow \boxed{w = \sum_{i=1}^N \alpha_i y_i . x_i}$$

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

Hint:

$$\left(\sum_{i=1}^N \alpha_i\right)\left(\sum_{i=1}^N \alpha_i\right) \neq \sum_{i=1}^N \alpha_i^2$$

$$\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j$$

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

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

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

$$= \cancel{\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \cancelto{0}{\sum_{i=1}^{N} \alpha_i y_i} w_0 + \sum_{i=1}^{N} \alpha_i$$

DUAL

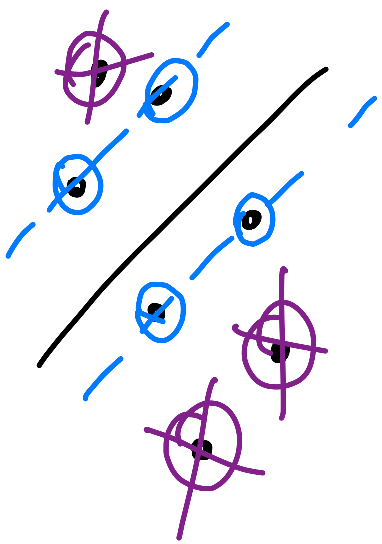
maximize $\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j$

subject to: $\sum_{i=1}^{N} \alpha_i y_i = 0$

$\alpha_i \geq 0 \quad \forall i$

\# of constraints = 1

\# of decision variables = N



∘ $\alpha_i > 0 \Rightarrow$ support vectors

∘ $\alpha_i = 0$