Kernel Machines

different models (inductive bias) $\Rightarrow$ different assumptions
$\Rightarrow$ 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)

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

representer theorem
$\Downarrow$

$$\underbrace{w}_{D\times 1} = \sum_{i=1}^{N} \underbrace{\alpha_i}_{1\times 1} . \underbrace{x_i}_{D\times 1}$$

α_i ↳ is mostly zero

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

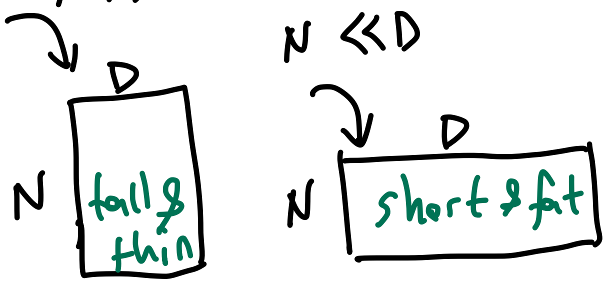
\# of parameters $= D+1$

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

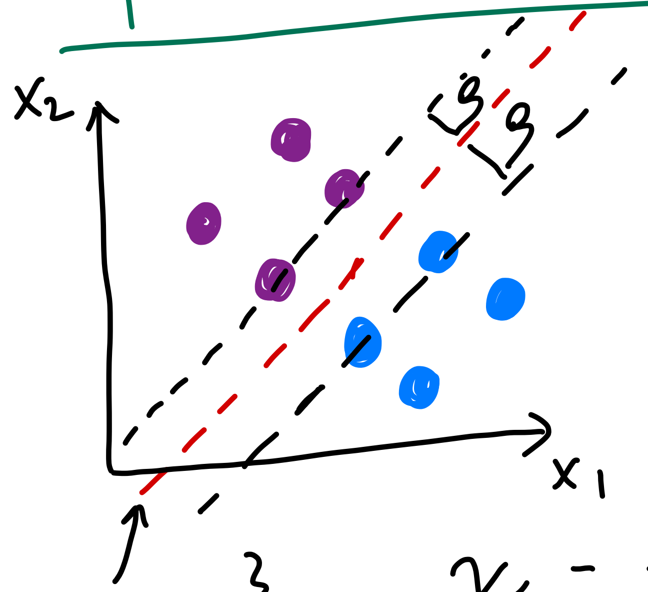
$\theta = \{\alpha_1, \alpha_2, \ldots, \alpha_N, w_0\}$

\# of parameters $= N+1$

$N \gg D$ $\qquad$ $N \ll D$



Optimal Separating Hyperplane



$+$ (purple), $-$ (blue)

ρ : margin

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

$w^T . x + w_0 = 0$

$$\left.\begin{aligned} 3x + 4y - 5 &= 0 \\ -3x - 4y + 5 &= 0 \\ 6x + 8y - 10 &= 0 \end{aligned}\right\} \text{equivalent}$$

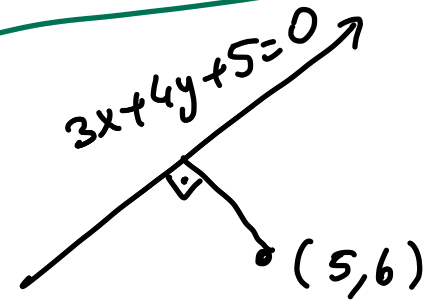
$$\boxed{\begin{aligned} -1(x+y) &\geq (-5) - 1 \\ -x - y &\leq +5 \end{aligned}}$$

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

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

$$\Downarrow$$

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



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

$$\frac{|w^T . x_i + w_0|}{\|w\|_2} = \frac{y_i(w^T . x_i + w_0)}{\|w\|_2} \geq \rho \Rightarrow y_i(w^T x_i + w_0) \geq \underbrace{\|w\|_2 . \rho}_{1}$$

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

minimize $x^2-6x+10$

$-x^2+8x-14$

$-16+32-16$

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

$x^*=3$

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

$\boxed{\text{maximize } -x^2+8x-14}$

$\boxed{\begin{array}{l}\text{minimize } x^2-6x+10 \\ \text{subject to: } x \geq 4\end{array}}$

$\Rightarrow$ minimize $x^2-6x+10$

subject to $x-4 \geq 0$] λ

minimize this function

Lagrangian$(x,\lambda) = [x^2-6x+10] - \lambda \cdot [x-4]$

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

$\lambda = 2x-6$

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

$= -x^2+8x-14$ ⇐ maximize

$x^*=4$

$$\left[\|w\|_2 \cdot g = 1\right] \quad \text{maximize } g \iff \text{minimize } \|w\|_2$$

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

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

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

$$\text{subject to: } \quad y_i(\underset{\checkmark}{w}^{\overset{?}{T}} \cdot \underset{\checkmark}{x_i} + \underset{?}{w_0}) \geq 1 \quad \forall i \quad \alpha_i$$

$D+1$ decision variables

N constraints

$$\|w\|_2^2 = w^T \cdot w$$

$$L_p = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{N} \alpha_i \left[y_i(w^T \cdot x_i + w_0) - 1\right].$$

$$\frac{\partial L_p}{\partial w} = \frac{\partial\left[\frac{1}{2} \cdot w^T \cdot w - \sum_{i=1}^{N} \alpha_i \left[y_i(w^T \cdot x_i + w_0) - 1\right]\right]}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

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

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

$$\cancel{\sum_{i=1}^{N} \alpha_i^2 y_i^2 x_i^T \cdot x_i}$$

$$(a_1 + a_2)(a_1 + a_2) \neq (a_1 + a_2)^2$$

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

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

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

$$L_p = \overbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T . x_j}^{\frac{1}{2}\|w\|_2^2} - \sum_{i=1}^{N} \alpha_i \left[y_i \left[\left[\sum_{j=1}^{N} \alpha_j y_j x_j\right]^T . x_i + w_0 \right] - 1 \right]$$

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

$$0$$

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

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

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

$\overset{?}{\alpha_i} \geq 0$

$$\boxed{x_i^T . x_j = x_j^T . x_i}$$

$\#$ of decision variables $= N$

$\#$ of constraints $= 1$