Kernel Machines:

different models ⇒ different assumptions, different objective functions $\Big\}$ inductive bias

support vector machines (SVM)

↳ they do not care about probabilities or densities.

↳ weights (w) can be written in terms of training points.

$$g(x) = \underset{D\times 1}{\underline{\underline{w}}}^T . x + \underset{1\times 1}{\underline{\underline{w_0}}} \qquad \underbrace{x_1, x_2, \ldots, x_D}_{\text{D-dimensional inputs}}$$

(D+1) parameters

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iD} \end{bmatrix}$$

$$\underset{D\times 1}{W} = \sum_{i=1}^{N} \underset{\text{scalars (sparse)}}{a_i} \underset{D\times 1}{x_i}$$

representer theorem

↳ discriminant function can be written in terms of dot products between x_i & $\underline{\underline{x}}$

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

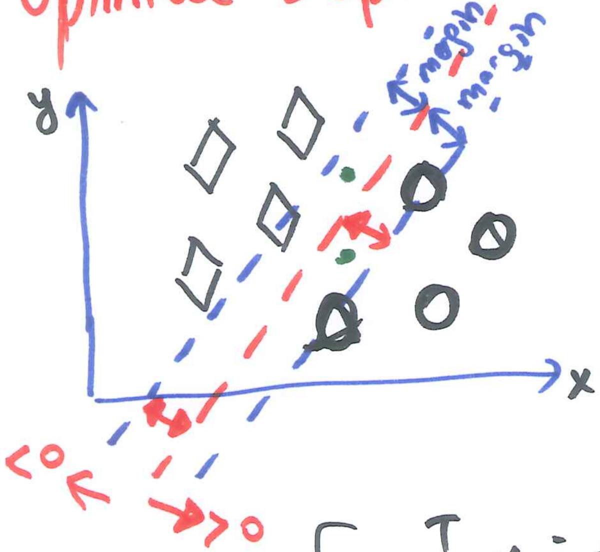
$$\boxed{= \sum_{i=1}^{N} a_i x_i^T . x + w_0}$$

↳ convex problem ⇒ global optimum

$$\theta = \{a_1, a_2, \ldots, a_N, w_0\}$$

if $N \ll D$

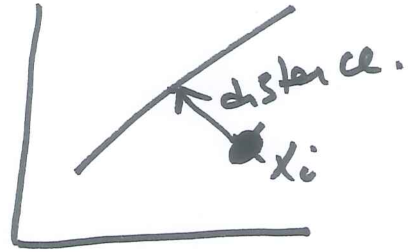
Optimal Separating Hyperplane



◇ = ● −

O = +

$$2x + 3y + 5 = 0$$
$$4x + 6y + 10 = 0$$
$$-2x - 3y - 5 = 0$$



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

$$\begin{cases} w^T . x_i + w_0 \geq 1 & \text{if } y_i = +1 \\ w^T . x_i + w_0 \leq -1 & \text{if } y_i = -1 \end{cases}$$

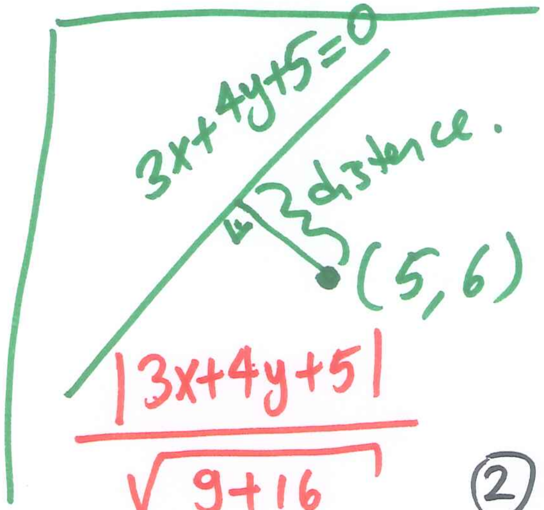
$$\boxed{y_i (w^T . x_i + w_0) \geq 1} \quad \forall i$$

$$-1 (w^T . x_i + w_0) \geq 1.$$
$$w^T . x_i + w_0 \leq -1$$

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

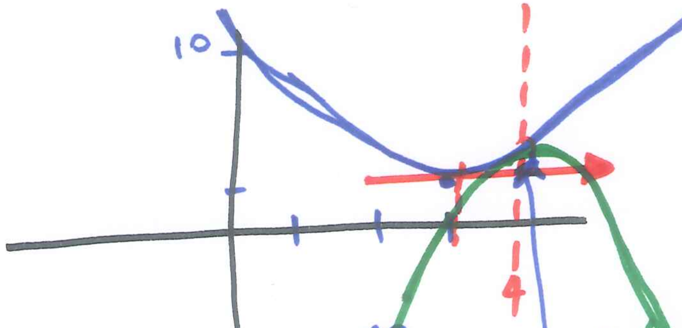
to find a unique solution $\Rightarrow ||w|| . \rho = 1$

$$\boxed{y_i (w^T . x_i + w_0) \geq 1}$$



$$\frac{|3x + 4y + 5|}{\sqrt{9 + 16}}$$

minimize $x^2-6x+10 \Rightarrow x^*=3$



minimize $x^2-6x+10$

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

$x-4 \geq 0$

Lagrangian $= x^2-6x+10-\lambda(x-4)$

$$\frac{\partial \text{Lagrangian}}{\partial x} = 2x-6-\lambda \Rightarrow \lambda = \boxed{2x-6}$$

$$x^2-6x+10-(2x-6)(x-4)$$

$$x^2-6x+10-2x^2+14x-24$$

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

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

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

$-x^2+8x-14$

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

$\boxed{x^*=4}$

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

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

PRIMAL SVM PROBLEM

$\alpha_i \geq 0$

$$\|w\|.g = 1.$$

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

$$\|w\|^2 = w^T.w$$

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

$$\frac{\partial L_P}{\partial w} = \cancel{\frac{1}{2}} \cancel{2}.w - \sum_{i=1}^{N} \alpha_i y_i . x_i = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N} \alpha_i . y_i . x_i$$

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

$$L_D = \cancel{\frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)^T \left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)} - \underbrace{\frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)^T . \left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)}_{\sum_{i=1}^{N} \alpha_i y_i w^T . x_i} - \cancel{\sum_{i=1}^{N} \alpha_i y_i w_0} + \sum_{i=1}^{N} \alpha_i$$

$$\cancel{w_0\left(\sum \alpha_i y_i\right)}$$

$$\boxed{L_D = -\frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)^T \left(\sum_{i=1}^{N} \alpha_i y_i x_i\right) + \sum_{i=1}^{N} \alpha_i}$$

$(x_1+x_2+x_3)(x_1+x_2+x_3) \rightarrow x_1^2+x_2^2+x_3^2+\ldots\ldots$

$$\left(\sum_{i=1}^{3} x_i\right)\left(\sum_{i=1}^{3} x_i\right) = \boxed{\sum_{i=1}^{3}\sum_{j=1}^{3} x_i \cdot x_j}$$

$$\sum_{i=1}^{3} x_i^2 = x_1^2+x_2^2+x_3^2$$

$$L_D = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \overset{U}{\alpha_i} \cdot \overset{U}{\alpha_j} \overset{K}{y_i} \overset{K}{y_j} \boxed{\overset{K\,T}{x_i} \cdot \overset{K}{x_j}} + \sum_{i=1}^{N} \overset{U}{\alpha_i}$$

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

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

$\alpha_i \geq 0$

DUAL SVM PROBLEM

DUAL SVM PROBLEM

# of decision variables : N

# of constraints : 1

PRIMAL SVM PROBLEM

# of decision variables : D+1

# of constraints : N