PRIMAL PROBLEM

$w_1^2 + w_2^2 + \cdots + w_D^2$

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

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

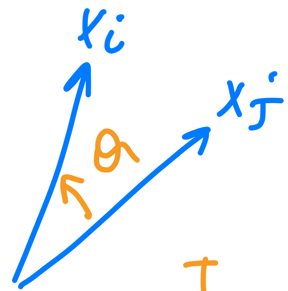
for all

subject to : $y_i(w^T . x_i + w_0) \geq 1 \quad \forall i$ $\Rightarrow$ separation constraints

Decision variables = $\{w, w_0\}$ $\quad$ # of decision variables = $D+1$

# of constraints = N

DUAL PROBLEM



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

subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ $\Rightarrow$ only constraint

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

$\cos \theta = \frac{x_i^T . x_j}{\|x_i\|_2 \|x_j\|_2}$

Decision variables = $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ $\quad$ # of decision variables = N

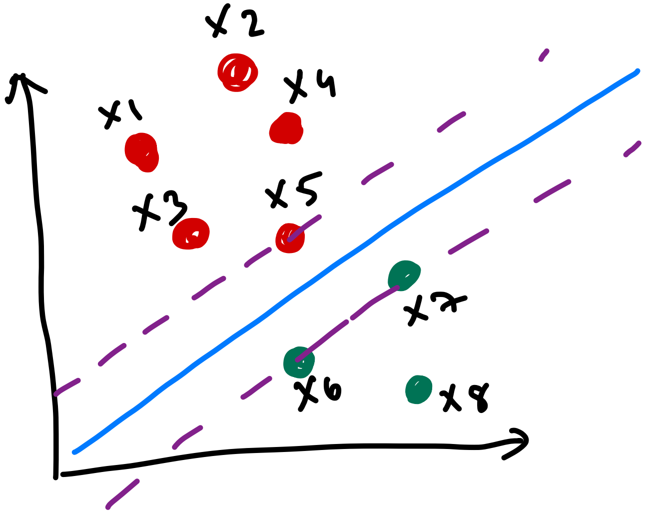
# of constraints = 1

let us assume we solved the dual problem $\Rightarrow \alpha^*$

$$W^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

most of α_i^*'s are zero

if $\alpha_i^* > 0$, x_i is called a "support vector"



$\alpha_1^* = 0$ $\alpha_5^* \geq 0$

$\alpha_2^* = 0$ $\alpha_6^* \geq 0$

$\alpha_3^* = 0$ $\alpha_7^* \geq 0$

$\alpha_4^* = 0$ $\alpha_8^* = 0$

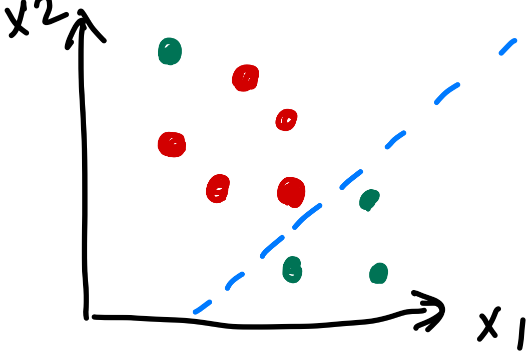
test data point

$$g(x) = W^T \cdot x + w_0 = \left(\sum_{i=1}^{N} \alpha_i^* y_i x_i\right)^T \cdot x + w_0$$

$$= \sum_{i=1}^{N} \underline{\alpha_i^*} y_i \underline{x_i^T \cdot x} + w_0$$

We do not have to store x_1, x_2, x_3, x_4, x_8 in the memory!

Nonseparable Case



$$\text{minimize} \quad \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{N} \varepsilon_i$$

$C \to +\infty$

Subject to:

$\alpha_i \quad \left[y_i(w^T . x_i + w_0) \geq 1 - \varepsilon_i \quad \forall i \right.$

$\beta_i \quad \left[\varepsilon_i \geq 0 \quad \forall i \right.$

$\varepsilon_i \to$ xsi

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

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

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

$$\frac{\partial L_p}{\partial \varepsilon_i} = C - \alpha_i - \beta_i = 0 \quad \forall i \Rightarrow \alpha_i + \beta_i = C \Rightarrow 0 \leq \alpha_i \leq C \quad \forall i$$

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

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

$\boxed{C \geq} \alpha_i \geq 0 \quad \forall i$

Kernel Trick

$x \in \mathbb{R}^D \qquad z \in \mathbb{R}^Q \qquad$ usually $D \ll Q$

phi $\leftarrow \Phi : X \rightarrow Z$

mapping function

$\underline{D=1}$

x_i

$\underline{Q=3}$

$z_i = \begin{bmatrix} x_i^2 \\ x_i^3 \\ x_i \end{bmatrix}$

X domain

$W = \sum_{i=1}^{N} \alpha_i y_i . x_i$

$f(x) = W^T . X + w_0$

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

$\Rightarrow$

Z domain

$W = \sum_{i=1}^{N} \alpha_i y_i z_i = \sum_{i=1}^{N} \alpha_i y_i \Phi(x_i)$

$f(z) = W^T . z + w_0$

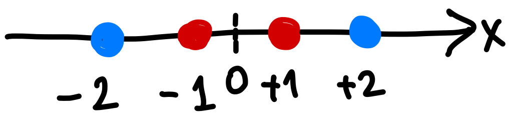
$= W^T . \Phi(x) + w_0$

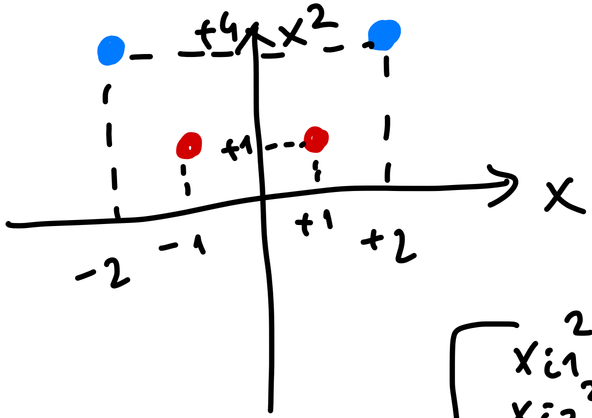
$= \sum_{i=1}^{N} \alpha_i y_i \boxed{\Phi(x_i)^T . \Phi(x)} + w_0$

$x_i^T . x_j \Rightarrow k(x_i, x_j)$

$x_i^T x \Rightarrow k(x_i, x)$

kernel function $\leftarrow k(x_i, x)$





$$\Phi(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \Rightarrow \Phi(x_i) = z_i = \begin{bmatrix} x_{i1}^2 \\ x_{i2}^2 \\ \sqrt{2}x_{i1}x_{i2} \\ \sqrt{2}x_{i1} \\ \sqrt{2}x_{i2} \\ 1 \end{bmatrix}$$

$D = 2$ $\qquad$ $d = 6$

$$\Phi(x_i)^T \cdot \Phi(x_j) = \begin{bmatrix} x_{i1}^2 & x_{i2}^2 & \sqrt{2}x_{i1}x_{i2} & \sqrt{2}x_{i1} & \sqrt{2}x_{i2} & 1 \end{bmatrix} \begin{bmatrix} x_{j1}^2 \\ x_{j2}^2 \\ \sqrt{2}x_{j1}x_{j2} \\ \sqrt{2}x_{j1} \\ \sqrt{2}x_{j2} \\ 1 \end{bmatrix}$$

$$= x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 1$$

$$= (\underbrace{x_{i1}x_{j1} + x_{i2}x_{j2}}_{\begin{bmatrix} x_{i1} & x_{i2} \end{bmatrix} \begin{bmatrix} x_{j1} \\ x_{j2} \end{bmatrix}} + 1)^2 = (x_i^T \cdot x_j + 1)^2$$

Linear Kernel : $k(x_i, x_j) = x_i^T \cdot x_j \Rightarrow \Phi(x_i) = x_i$

Polynomial Kernel : $k(x_i, x_j) = (x_i^T \cdot x_j + 1)^q$

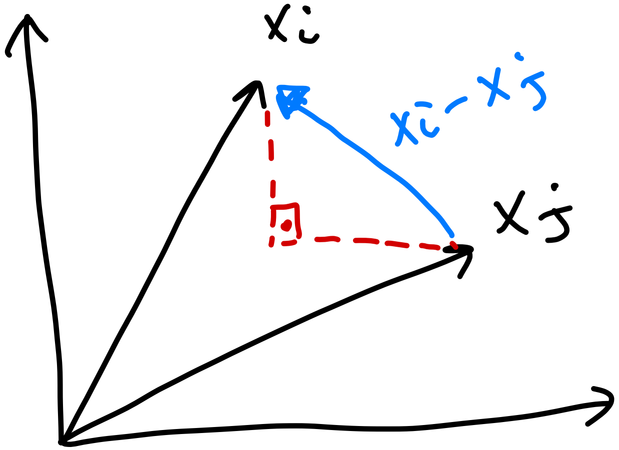
qth order polynomial kernel

hyperbolic tangent

Sigmoidal Kernel : $\tanh(2x_i^T \cdot x_j + 1)$

Gaussian Kernel : $\exp\left(-\frac{\|x_i - x_j\|_2^2}{2s^2}\right)$

∞th order polynomial



$$\|x_i - x_j\|_2^2 = \left(\cancel{\sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2}}\right)^{\cancel{2}}$$

Hademard multiplication

maximize $1^T.\alpha - \frac{1}{2}\alpha^T(K \circ (yy^T)).\alpha$

subject to: $y^T.\alpha = 0$

$c.1 \geq \alpha \geq 0$

$[\ ] \to [k(x_i,x_j)] \to [y_i y_j]\ [\ ]$

$[A] \circ [B] = [C]$

$C_{ij} = a_{ij} \cdot b_{ij}$

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

$$\sum_{i=1}^{N} \alpha_i y_i$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$