PRIMAL PROBLEM:

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

$y_i \in \{-1, +1\}$

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

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

Decision variables $= \{w, w_0\}$

# of decision variables $= D+1$

# of constraints $= N$

DUAL PROBLEM:

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

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

$\underline{\alpha_i \geq 0} \quad \forall i$

Decision variables $= \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$

# of decision variables $= N$

# of constraints $= 1$

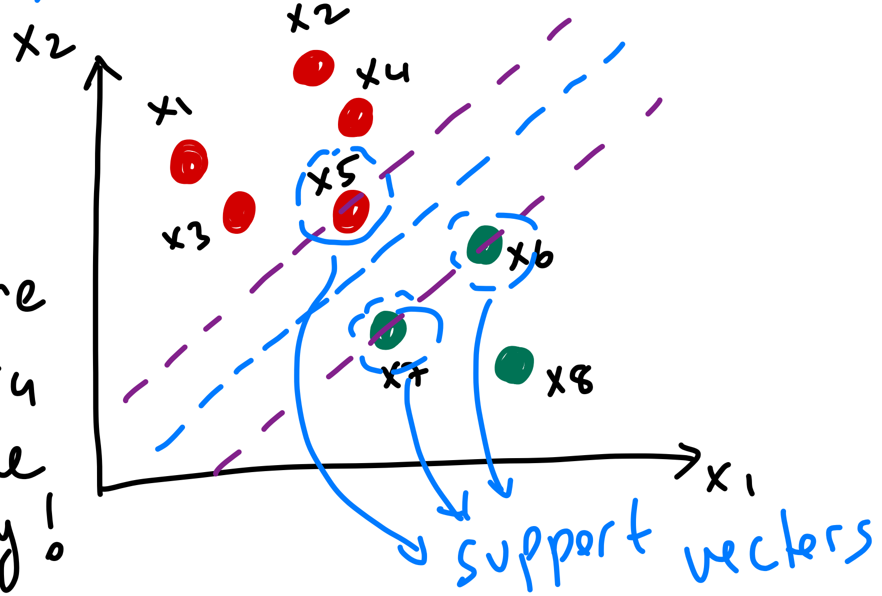
Let us assume we solved the dual problem $\Rightarrow \alpha^*$

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

$\hookrightarrow$ the solution to the primal problem

$\Big\}$ most of α_i^*'s are zero

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



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

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

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

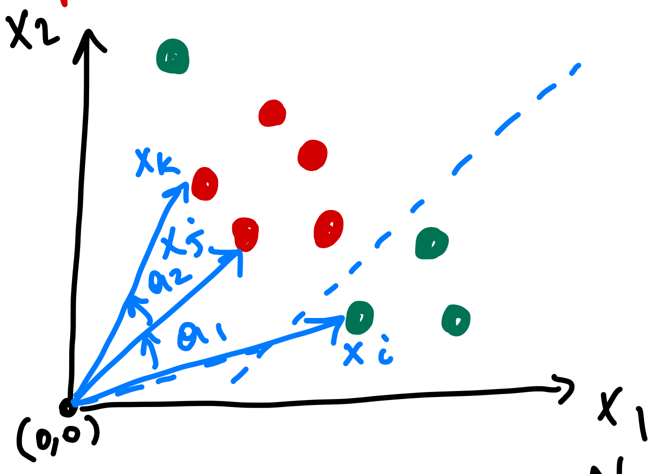
$\alpha_3^* = 0$ $\quad$ $\alpha_7^* \geq 0$

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

$$f(x) = w^T . x + w_0 = \left(\sum_{i=1}^{N} \alpha_i^* y_i . x_i \right)^T . x + w_0$$

$\hookrightarrow$ when we are given a test data point.

Nonseparable Case:



PRIMAL

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

subject to: $\alpha_i\,[\; y_i(w^T.x+w_0) \geq 1-\epsilon_i \quad \forall i$

$\beta_i\,[\; \epsilon_i \geq 0 \quad \forall i$

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

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

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

$$\frac{\partial L_p}{\partial \epsilon_i} = C - \alpha_i - \beta_i = 0 \Rightarrow \alpha_i + \beta_i = C \Rightarrow 0 \leq \alpha_i \leq C$$

DUAL

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

?

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

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

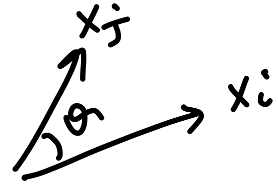
if you set $C=+\infty$, you do not allow slack variables.

Kernel Trick:

$x \in \mathbb{R}^D$ $\quad z \in \mathbb{R}^Q$

usually $Q >> D$

much much larger



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

$$\Phi : x \to z$$

mapping function

$D=1$

x_i

$Q=3$

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

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

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

$k(x_i, x) \Rightarrow$ similarity metric

$$\Rightarrow f(z) = w^T . z + w_0$$
$$= w^T . \Phi(x) + w_0$$
$$= \sum_{i=1}^{N} \alpha_i y_i \boxed{\underbrace{\Phi(x_i)^T}_{z_i^T} . \underbrace{\Phi(x)}_{z}} + w_0$$

$k(x_i, x)$

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

$k(x_i, x_j)$

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

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

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

$D = 2$

$n = 6$

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

$$k(x_i, x_j) = [x_i^T x_j + 1]^2$$

$\Downarrow$

Second-order polynomial kernel

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

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

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

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

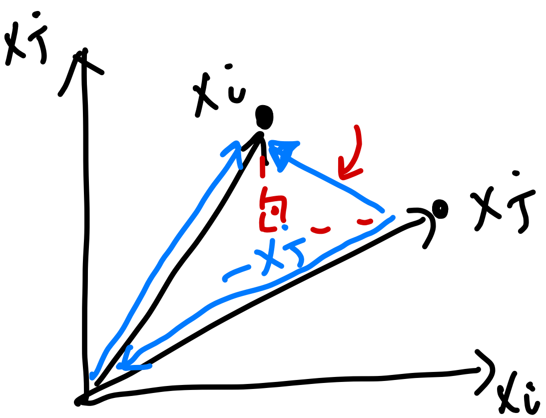
q th order polynomial kernel.

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

squared Euclidean distance.

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

∞th order polynomial



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

maximize $\displaystyle\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j \underbrace{y_i y_j \underbrace{k(x_i,x_j)}}$

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

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

$$\overbrace{[1\ 1 \cdots 1]}^{1^T}\underbrace{\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix}}_{\alpha}$$

$$\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix}$$

Hadamard product

maximize $\mathbf{1}^T\cdot\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\cdot\left(K \circ (yy^T)\right)\boldsymbol{\alpha}$

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

$C\cdot\mathbf{1} \geq \boldsymbol{\alpha} \geq \mathbf{0}$

$\mathbf{1}$ vector of size N

$\mathbf{0}$ vector of size N

$$\mathbf{K} = \begin{array}{c} \begin{bmatrix} & & | & & \\ & & | & & \\ - & - & - & - & - \\ & & | & & \\ & & | & & \end{bmatrix} \begin{array}{l} \\ \\ i\text{th row} \\ \\ \\ \end{array} \\ j\text{th column} \quad N\times N \end{array}$$

kernel matrix

$$yy^T = \begin{array}{c} \begin{bmatrix} & & | & & \\ - & - & y_i y_j & - & - \\ & & | & & \end{bmatrix} \begin{array}{l} \\ i\text{th row} \\ \\ \end{array} \\ j\text{th column} \quad N\times N \end{array}$$

$$k_{ij} = k(x_i, x_j)$$

$$K \circ (yy^T) = \begin{array}{c} \begin{bmatrix} - & y_i y_j k(x_i, x_j) & - \end{bmatrix}_{N \times N} \\ j^{th} \text{ column} \end{array} \; i^{th} \text{ row}$$

$$\alpha^T . (K \circ (yy^T)) . \alpha$$

$$\begin{array}{c} \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \cdots & \alpha_N \end{bmatrix} \\ i^{th} \text{ column} \end{array} \begin{array}{c} \begin{bmatrix} K \circ (yy^T) \end{bmatrix} \; i^{th} \text{ row} \\ j^{th} \text{ column} \end{array} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \; j^{th} \text{ row}$$

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

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

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

$$C.1 \geq \alpha \geq 0$$

Quadratic Programming (QP)

objective function should be concave with respect to α.

!!! K should be a positive semi-definite matrix to obtain a concave function.

$$\boxed{a^T.K.a \geq 0 \quad \forall a}$$