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

subject to: $y_i(\underbrace{w^T.x_i + w_0}_{\text{decision variables} \Rightarrow (D+1) \text{ variables}}) \geq 1 \quad \forall i$

PRIMAL PROBLEM

$\Downarrow$

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

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

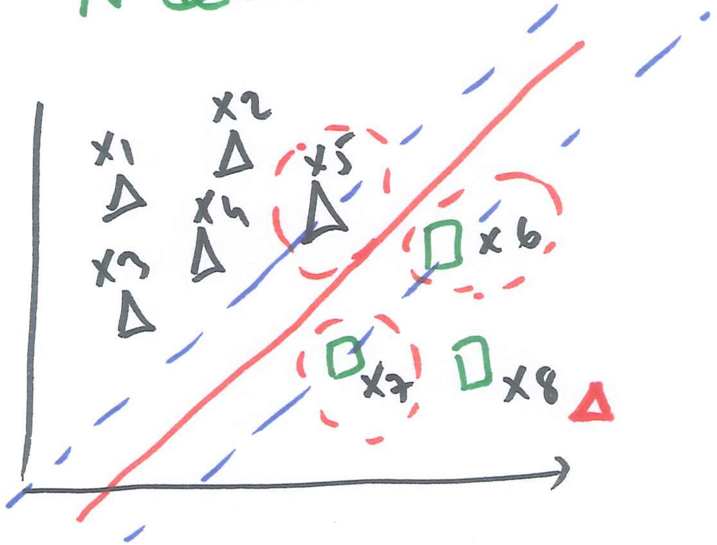
DUAL PROBLEM

N decision variables.

α^* solution of this problem

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

"support vector machine"



$\alpha_1 = 0$

$\alpha_2 = 0$

$\alpha_3 = 0$

$\alpha_4 = 0$

$\alpha_5 > 0$

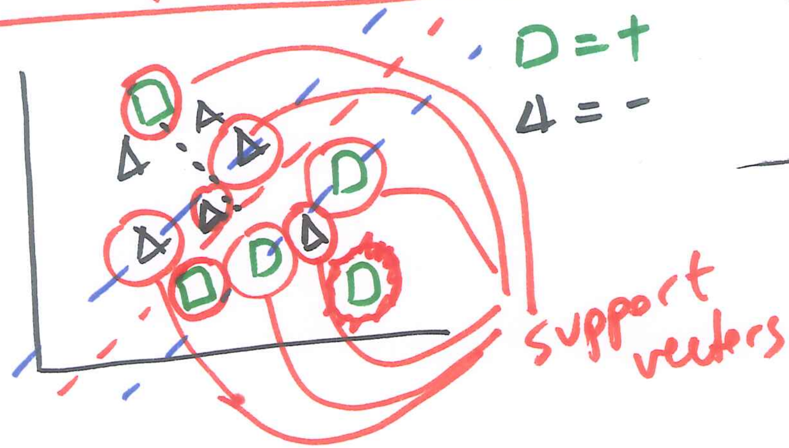
$\alpha_6 > 0$

$\alpha_7 > 0$

$\alpha_8 = 0$

support vectors

Nonseperable case:



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

$$\varepsilon_i \geq 0 \quad \forall i$$

model complexity

error term

penalty factor (>0)

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

subject to $y_i(w^T x_i + w_0) \geq 1 - \varepsilon_i \quad \forall i \quad \alpha_i$

$\varepsilon_i \geq 0 \quad \forall i \quad \beta_i$

$\alpha_i \geq 0 \qquad \beta_i \geq 0$

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

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

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

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

Exercise #1.

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

$\Rightarrow$ this is the only place where x_i's appear.

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

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

\# of decision variables: N

$$N \left\{ \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & x_1^T x_N \\ & \ddots & & \\ & & \ddots & \\ & & & x_N^T x_N \end{bmatrix} \right.$$

Kernel Trick

$$\underline{\Phi} : x \rightarrow z \qquad x \in \mathbb{R}^{\underline{D}} \qquad z \in \mathbb{R}^{\underline{Q}}$$

mapping functions.

usually $\overset{H}{Q} \gg D$

$$\underset{x}{\begin{array}{c} x \\ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{array}} \rightarrow \underset{z}{\begin{array}{ccc} x & x^2 & x^3 \\ \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \begin{bmatrix} 1 \\ 4 \end{bmatrix} & \begin{bmatrix} 1 \\ 8 \end{bmatrix} \end{array}}$$

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

$$f(x) = w^T \Phi(x) + w_0 \Rightarrow \sum \alpha_i y_i \underbrace{\boxed{\Phi(x_i)^T \Phi(x)}}_{x_i^T . x} + w_0$$

$$K_{ernel} = \begin{bmatrix} \Phi(x_1)^T \Phi(x_1) & \cdots & \Phi(x_1)^T \Phi(x_N) \\ & \ddots & \\ & & \Phi(x_N)^T . \Phi(x_N) \end{bmatrix}$$

$$\text{maximize} \quad \underline{\sum_{i=1}^{N} \alpha_i} - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \underbrace{\boxed{k(x_i, x_j)}}_{x_i^T x_j}$$

kernel trick.

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

$$C \geq \alpha_i \geq 0$$

$$\boxed{f(x) = \sum \alpha_i y_i k(x_i, x) + w_0}$$

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}_{2\times1} \Rightarrow \Phi(x_i) = \begin{bmatrix} x_{i1}^2 \\ x_{i2}^2 \\ \sqrt{2}\, x_{i1} . x_{i2} \\ \sqrt{2}\, x_{i1} \\ \sqrt{2}\, x_{i2} \\ 1 \end{bmatrix}_{6\times1}$$

$D=2 \Longrightarrow d=6$

$$\Big[\overbrace{x_{i1} x_{j1} + x_{i2} x_{j2}}^{x_i^T x_j} + 1\Big]^2$$

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

$\Downarrow$ second order polynomial kernel.

$$\Phi(x_i)^T . \Phi(x_j) = [x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}, 1] \begin{bmatrix} x_{j1}^2 \\ x_{j2}^2 \\ \sqrt{2}\, x_{j1} . x_{j2} \\ \sqrt{2}\, x_{j1} \\ \sqrt{2}\, x_{j2} \\ 1 \end{bmatrix} = \boxed{\begin{aligned} & x_{i1}^2 . x_{j1}^2 + x_{i2}^2 . x_{j2}^2 \\ & + 2x_{i1} x_{i2} x_{j1} x_{j2} \\ & + 2x_{i1} x_{j1} + 2x_{i2} x_{j2} \\ & + 1 . \end{aligned}}$$

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

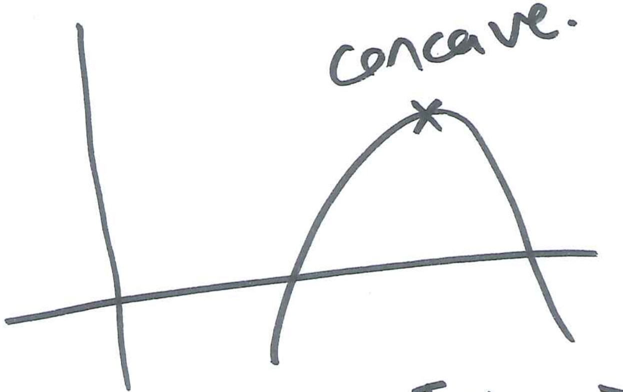
subject to: $\alpha^T.y = 0$, $\mathbf{1}.C \geq \alpha \geq 0$

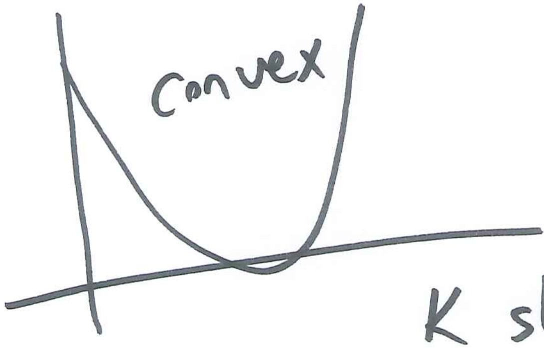
$$A \circ B = \begin{bmatrix} a_{11}.b_{11} & a_{12}.b_{12} & \cdots \\ & \ddots & \\ & & \ddots \end{bmatrix}$$

$.*$

$c_{ij} = a_{ij} * b_{ij}$

$K \circ yy^T \Rightarrow$





K should be p.s.d.

positive semidefinite.

$x^T.Kx \geq 0 \quad \forall x.$

$a^T.Ka \geq 0 \quad \forall a$

polynomial kernel: $[x_i^T x_j + 1]^q \rightarrow q^{th}$ order polynomial kernel.

linear kernel: $x_i^T x_j \Rightarrow \Phi(x_i) = x_i$

sigmoidal kernel: $\tanh(2 x_i^T x_j + 1)$

gaussian kernel: $\exp\left[-\frac{\|x_i - x_j\|^2}{2s^2}\right] \rightarrow \infty^{th}$ order polynomial.

$1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \ldots$

$$K = \begin{bmatrix} x_i^T x_j & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & \ddots \end{bmatrix} \Rightarrow \text{p.s.d.}$$

$a > 0 \quad \underline{a\,k(x_i, x_j)} \rightarrow$ p.s.d.

$x^T . K . x \geq 0 \quad \forall x.$

$x^T . (a . K) x \overset{?}{\geq} 0$

$a . (x^T . K . x) \geq 0$

$\underset{>0}{a} . \underset{\geq 0}{(x^T . K . x)}$

$k_1(x_i, x_j) + k_2(x_i, x_j)$

$x^T . (K_1 + K_2) x \overset{?}{\geq} 0$

$\underset{\geq 0}{x^T . K_1 x} + \underset{\geq 0}{x^T . K_2 x} \geq 0$

$k_1(x_i, x_j) * k_2(x_i, x_j)$

$x^T (K_1 \circ K_2) x \overset{?}{\geq} 0$ Exercise #1?

Multiple kernel Learning (MKL)

$$k^*(x_i, x_j) = \sum_{m=1}^{P} \eta_m k_m(x_i, x_j) \quad \text{where } \eta_m \geq 0$$

$\eta_m \rightarrow$ kernel weights.

optimize α_i's and η_m's at the same time.

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i k^*(x_i, x) + w_0$$

$$= \sum_{i=1}^{N} \sum_{m=1}^{P} \underline{\underline{\alpha_i}}\, y_i\, \underline{\underline{\eta_m}}\, k_m(x_i, x) + w_0$$