PRIMAL PROBLEM:

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

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

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

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

\# of decision variables $= D+1$

\# of constraints $= N$

DUAL PROBLEM:

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

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

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

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

\# of decision variables $= N$

\# of constraints $= 1$

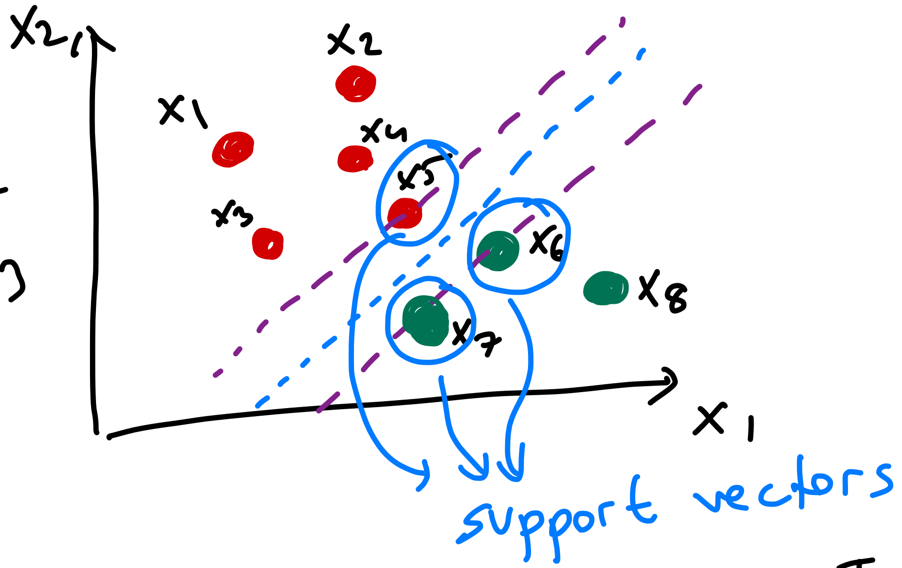
Let us assume we solved the dual problem $\Rightarrow \alpha^*$

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

$\}$ most of α_i^*'s are zero

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

↳ the solution to the primal problem



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

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

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

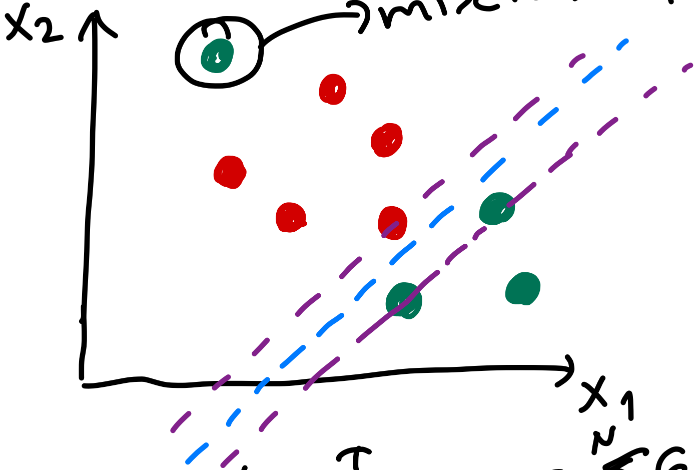
$\alpha_4^* = 0$ $\quad$ $\alpha_8^* = 0$

we do not have to store x_1, x_2, x_3, x_4, x_8 in the memory.

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

when we are given a test data point x

Nonseperable Case:



Primal

$$\left[\begin{array}{ll} \text{minimize} & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\epsilon_i \\ \text{subject to:} \quad \alpha_i: & y_i[w^T.x + w_0] \geq 1 - \epsilon_i \quad \forall i \\ \beta_i: & \epsilon_i \geq 0 \quad \forall i \end{array}\right.$$

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

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

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

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

Dual

$$\left[\begin{array}{ll} \text{maximize} & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \boxed{x_i^T x_j} \\ \text{subject to:} & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & \boxed{C \geq}\ \alpha_i \geq 0 \quad \forall i \end{array}\right.$$

only place we have x_i's

different from the previous

if you set $C = \infty$, you don't allow misclassified data points

Kernel Trick:

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

$\Phi : X \to Z$

$\hookrightarrow$ mapping function

$D = 1$ $\quad Q = 3$

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

$$z_i = \Phi(x_i)$$

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

$$f(x) = w^T . x + w_0 \qquad \Rightarrow \qquad f(z) = w^T . z + w_0$$

$$= \sum_{i=1}^{N} \alpha_i y_i \underbrace{\boxed{x_i^T . x}}_{k(x_i, x)} + w_0 \qquad = w^T . \Phi(x) + w_0$$

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

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

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

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

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

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

$D = 2$

$\theta = 6$

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

$$x_i^T . x_j \Rightarrow \Phi(x_i)^T \Phi(x_j)$$

$$\underbrace{k(x_i, x_j) = [x_i^T . x_j + 1]^2}_{\text{second order polynomial kernel}} = x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2 x_{i1} x_{i2} x_{j1} x_{j2} + 2 x_{i1} x_{j1} + 2 x_{i2} x_{j2} + 1$$

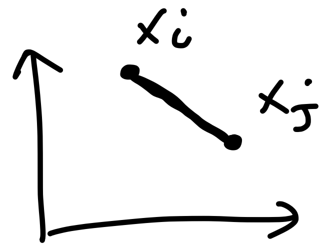
$$= (\underbrace{x_{i1} x_{j1} + x_{i2} x_{j2}}_{[x_{i1} \;\; x_{i2}] \begin{bmatrix} x_{j1} \\ x_{j2} \end{bmatrix} = x_i^T . x_j} + 1)^2 = [x_i^T x_j + 1]^2$$

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

Polynomial Kernel : $(x_i^T . x_j + 1)^q \rightarrow q^{th}$ order polynomial kernel

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

Gaussian Kernel : $\underbrace{\exp\left[- \frac{\overbrace{\|x_i - x_j\|_2^2}^{\text{squared Euclidean distance.}}}{2 . s^2} \right]}_{\infty^{th} \text{ order polynomial}}$



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

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

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

$$\text{maximize} \quad \underbrace{\mathbb{1}^T}_{1 \times N} \underbrace{\alpha}_{N \times 1} - \frac{1}{2} \underbrace{\alpha^T}_{1 \times N} \underbrace{(K \circ y y^T)}_{N \times N} \cdot \underbrace{\alpha}_{N \times 1}$$

$\circ$ → Hadamard product (entrywise multiplication)

$$\text{subject to:} \quad y^T \cdot \alpha = 0$$

$$0 \leq \alpha \leq C \cdot \mathbb{1}$$

$$\mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad 0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \sum_{i=1}^{N} \alpha_i y_i$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \sum_{i=1}^{N} \alpha_i$$

$$K = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & \boxed{k(x_1,x_N)} \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_N) \\ \vdots & \vdots & \ddots & \\ \boxed{k(x_N,x_1)} & k(x_N,x_2) & \cdots & k(x_N,x_N) \end{bmatrix}_{N \times N}$$

$$yy^T = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix} = \begin{bmatrix} y_1 \cdot y_1 & y_1 y_2 & \cdots & y_1 y_N \\ y_2 \cdot y_1 & y_2 y_2 & \cdots & y_2 y_N \\ \vdots & \vdots & \ddots & \vdots \\ y_N y_1 & y_N y_2 & \cdots & y_N y_N \end{bmatrix}_{N \times N}$$

$$K \circ yy^T = \begin{bmatrix} \cdots & y_i y_j k(x_i, x_j) & \cdots \end{bmatrix}$$

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

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \end{bmatrix} \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j)}_{\alpha^T \cdot (K \circ (yy^T)) \alpha}$$

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

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

$0 \leq \alpha \leq C.1$

concave with respect to α.

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

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

$\Rightarrow$ all eigenvalues of K should be nonnegative.

Constructing Kernels:

$k(x_i, x_j) \Rightarrow c.k(x_i, x_j)$

$\downarrow$ positive scalar

$a^T.K.a \geq 0 \Rightarrow a^T.(cK).a \geq 0$

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

$a^T K_1 a \geq 0$
$a^T K_2 a \geq 0$ $\Rightarrow a^T(K_1 + K_2).a \geq 0$

Exercise: $k_1(x_i, x_j)\,k_2(x_i, x_j)$ is also a valid kernel.