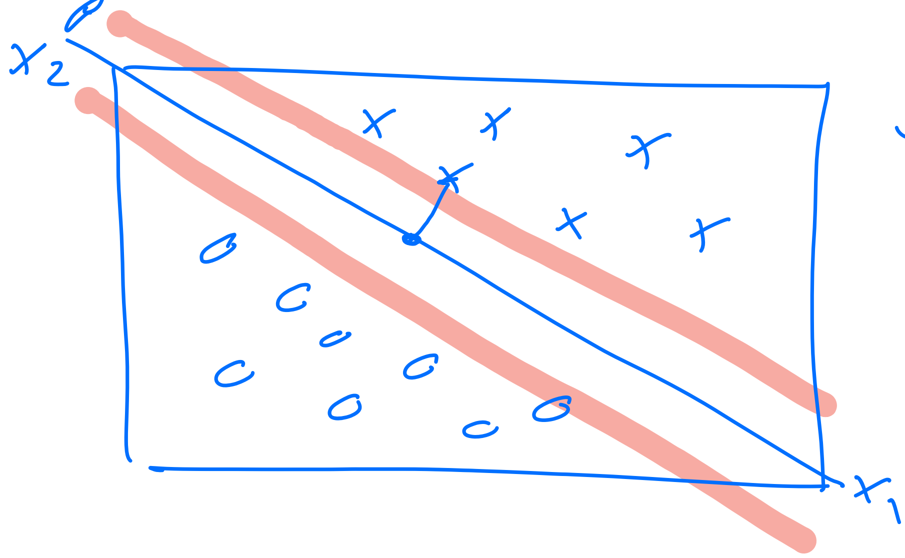
# Lecture December 13

## Support vector machines

$$f(x) = w^T x + b \approx y(x)$$




$y = \{-1, 1\}$

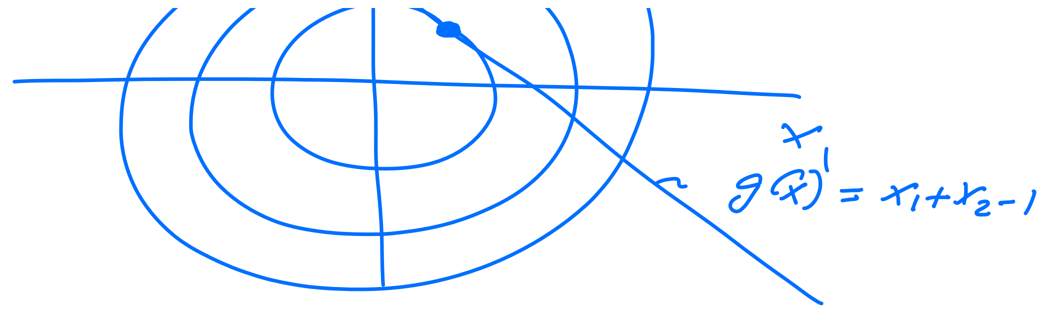
## Lagrange formalism

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$$

Example $\quad f(x) = 1 - x_1^2 - x_2^2$

$$g(x) = x_1 + x_2 - 1 = 0$$

when $g(x) > 0$, then constraint is inactive




$\nabla f(x) = 0$ , $g(x) > 0$ $\quad \lambda = 0$

$g(x) = 0$ , $\lambda \neq 0$

Maximize/minimize $f(x)$

subject to $g(x) \geq 0$

$$g(x) \geq 0$$
$$\lambda \geq 0$$
$$\lambda \cdot g(x) = 0$$

Karush-Kuhn-Tucker conditions

For SVM

$$\mathcal{L}(x, w, b, \lambda) = \frac{1}{2} w^T w - \sum_{i=0}^{n-1} \lambda_i \left( y_i (w^T x_i + b) - 1 \right)$$

$$g(x_i) = y_i \big( \underbrace{w^T x_i + b}_{f(x_i)} \big) - 1$$

$$\frac{\partial \mathcal{L}}{\;} = 0 = - \sum \lambda_i y_i$$

$\partial b$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 = w - \sum_i \lambda_i y_i x_i$$

$$\mathcal{L} = \sum_i \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$\lambda_i > 0 \quad \lambda_i \left[ y_i (w^T x_i + b) - 1 \right]$

outside the margin

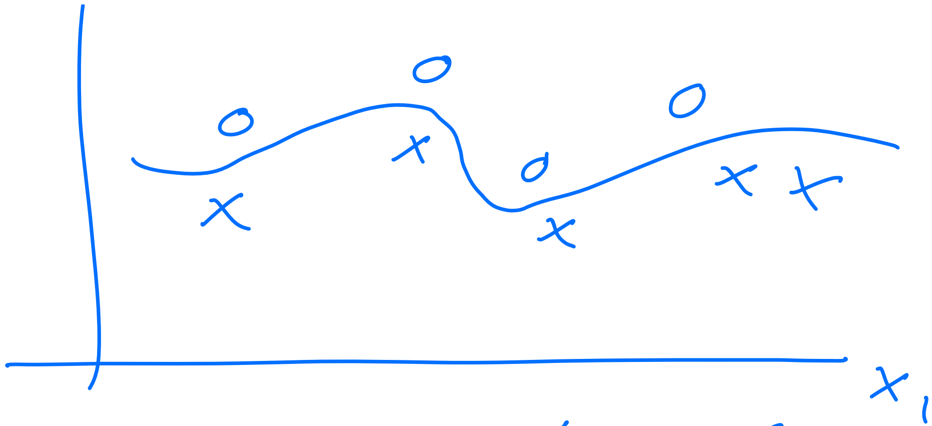
$$y_i (w^T x_i + b) > 1$$

$$\lambda^T = [\lambda_0, \dots \lambda_{n-1}]$$

$$\mathcal{L} = \lambda - \frac{1}{2} \lambda^T K \lambda$$

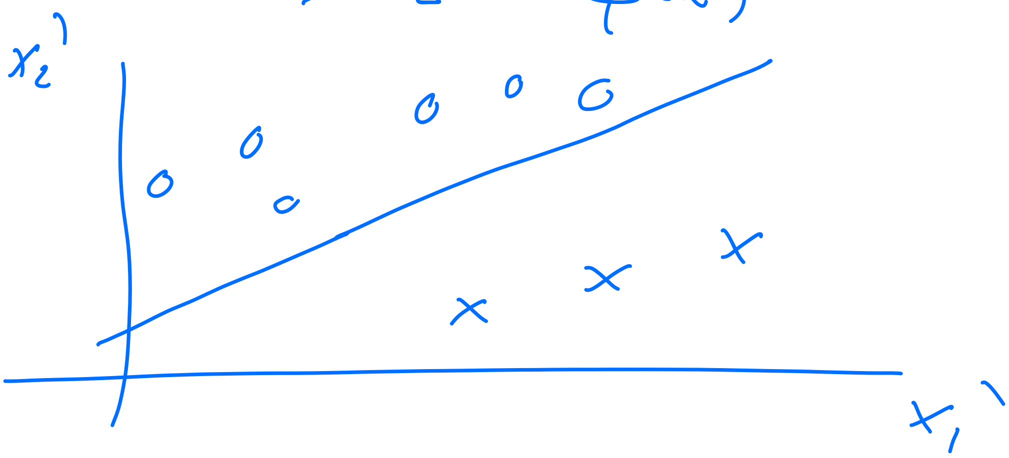
$$K = \begin{bmatrix} y_0 y_0 x_0^T x_0 & y_0 y_1 x_0^T x_1 & \dots & y_0 y_{n-1} x_0^T x_{n-1} \\ \vdots & \vdots & & \\ y_{n-1} y_0 x_{n-1}^T x_0 & \dots & \dots & y_{n-1} y_{n-1} x_{n-1}^T x_{n-1} \end{bmatrix}$$

$x_{21}$




smart transformation

$$x \rightarrow z = \phi(x)$$




$$f(x) = w^T x + b \quad \rightarrow$$

$$\tilde{f}(z) = w^T z + b$$

$$w = \sum_i \lambda_i y_i x_i \rightarrow \sum_i \lambda_i y_i z_i$$

$$\tilde{f}(z) = \left( \sum_i \lambda_i y_i z_i \right)^T z + b \quad \Rightarrow$$

$$= \left( \sum_i \lambda_i y_i \phi(x_i) \right)^T \phi(z) + b$$

$$\phi(x) = (1, x_1, x_1 x_2, x_2, x_1^2, x_2^2)$$

Example of a degree 2 polynomial

$$\phi(x)^T \phi(x') = z^T z' = F(x, x')$$

= kernel.

Example kernels:

- polynomials

  $$F(x,y) = (1 + x^T y)^s$$

  $s = 1 \Rightarrow$ linear kernel

- Sigmoid / tanh

  $$F(x,y) = \tanh(k x^T y - \delta)$$

- radial basis expansion

  $$F(x,y) = \exp\left\{ -(x-y)^2 / 2\sigma^2 \right\}$$