Support vector machines.

⇓

model-based learning.

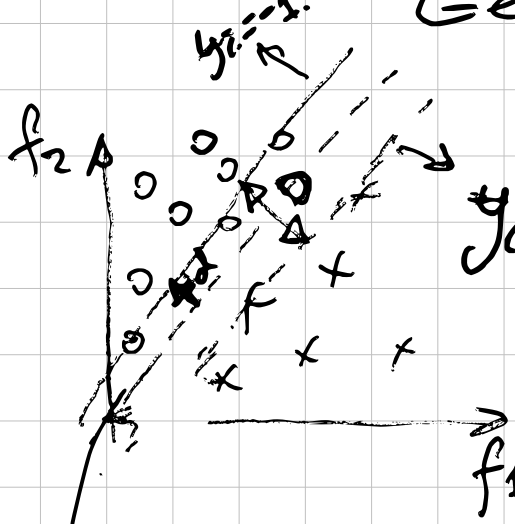
⇓

$\mathbf{X}_i = \{x_1, \ldots x_n\}$

$\mathbf{w} = \{w_1, \ldots, w_n\}$

$\sum w\mathbf{x} + b = \text{SCORE}$

Geometric approach



model

hyperplane in general

d is the distance from model to line delineating a class

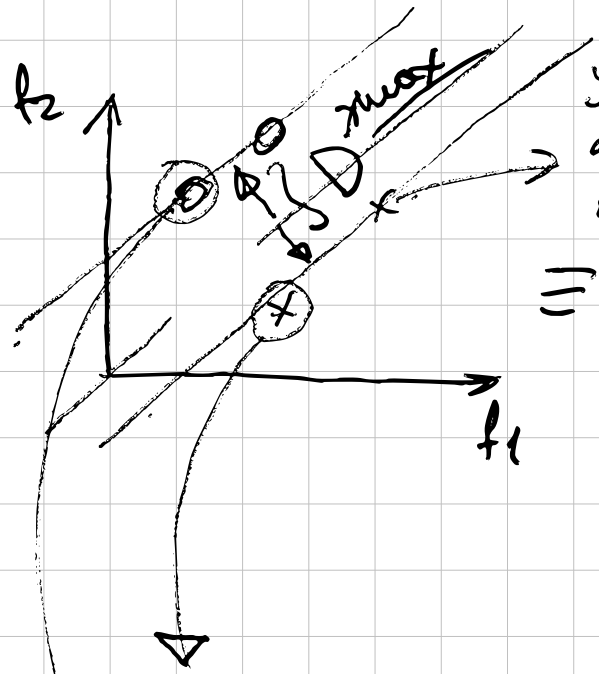
Illustration: 2 features, 2 classes

binary classific.

positive class ⇒ $y_i = +1$

negative class ⇒ $y_2 = -1$.

Data sample
Data point
≡ VECTOR of features



$y_i = +1$

data samples on borders of classes

≡ support vectors

$$w x_i + b \geq +1 \quad \text{if} \quad y_i = +1$$

$$w x_i + b \leq -1 \quad \text{if} \quad y_i = -1$$

b).

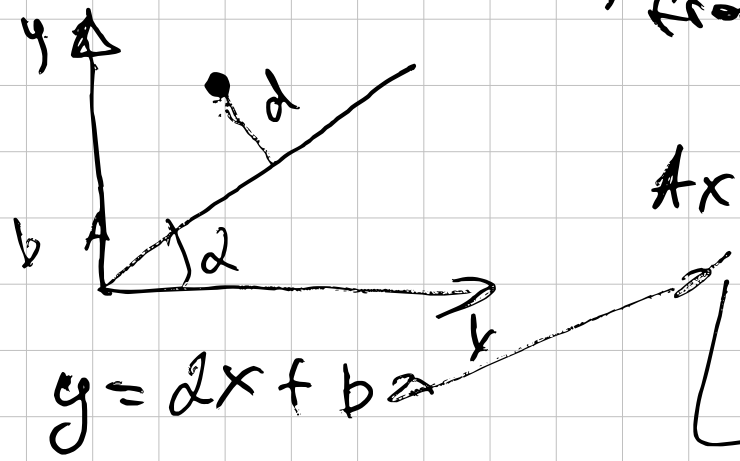
$$y_i \cdot (w x_i + b) \geq 1$$

We want to max $D = 2 \times d$

where d is dist. from model to support hyper plane

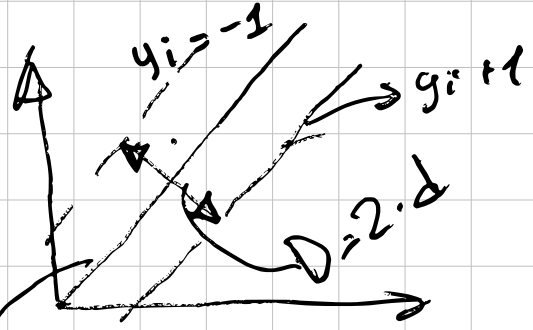
In 2D case hyper plane is line

from basic geometry



$$Ax + By + C = 0$$

$$d = \frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$



$$d = \frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$

$$d = \frac{|wx + b|}{\sqrt{\sum w_i^2}} = \frac{1}{\|w\|}$$

model ($wx + b$)

vector (matrix notation) $\sum w_i x_i + b$.

SVM: Want to "maximize"

$$f() = 2 \cdot d = \frac{2}{\|w\|}$$

minimize : $\frac{1}{f()} = \frac{1}{2}\|w\|$

$f^* = f()$

To make f^* convex : $\frac{1}{2}\|w\|^2$

Primary optimization formulation

$$f^* = \frac{1}{2}\|w\|^2$$

Constraint:

$$y_i(\underline{w}\underline{x}_i + \underline{b}) = 1$$

$$g: \quad y_i(w x_i + b) - 1 = 0$$

Solve optimization problem using Lagrangian method

$$\Rightarrow L = f^*(x) - \lambda g(x) \qquad + \sum \lambda_i y_i$$

$$= \frac{1}{2}\|w\|^2 - \sum \lambda_i [y_i(x_i w + b) - 1]$$

$w \cdot w$ dot prod.

$$\frac{\partial L}{\partial w} = \phi \quad : \quad w - \sum \alpha_i y_i x_i = 0$$

$$\frac{\partial L}{\partial b} = \phi \qquad -\sum \alpha_i y_i = \phi$$

$$w_i = \sum \alpha_i y_i x_i$$

Formulate secondary optimization problem — # of samples

$$L^* = \sum \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j (x_i \cdot x_j)$$

K

When dual problem is solved

few $\alpha_i > 0$

$f = wx + b$

(w) $= \sum \alpha_i y_i x_i$

(b) $= y_i - w x_i$

Multiclass classif.

- one vs. one: $\frac{n(n-1)}{2}$ SVMs
- one vs. all