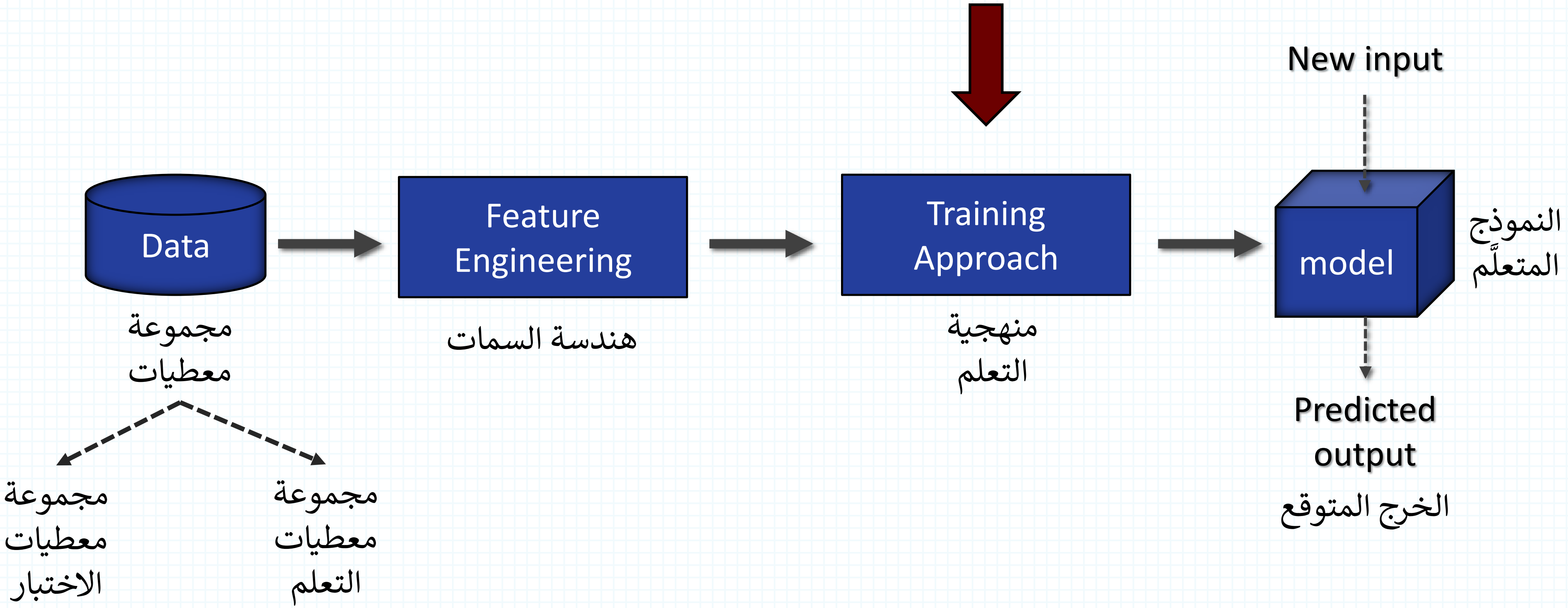
الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

مقرر تعلم الآلة | كلية الهندسة المعلوماتية | المحاضرة التاسعة

Support Vector Machine (SVM) 2

د. رياض سنبل

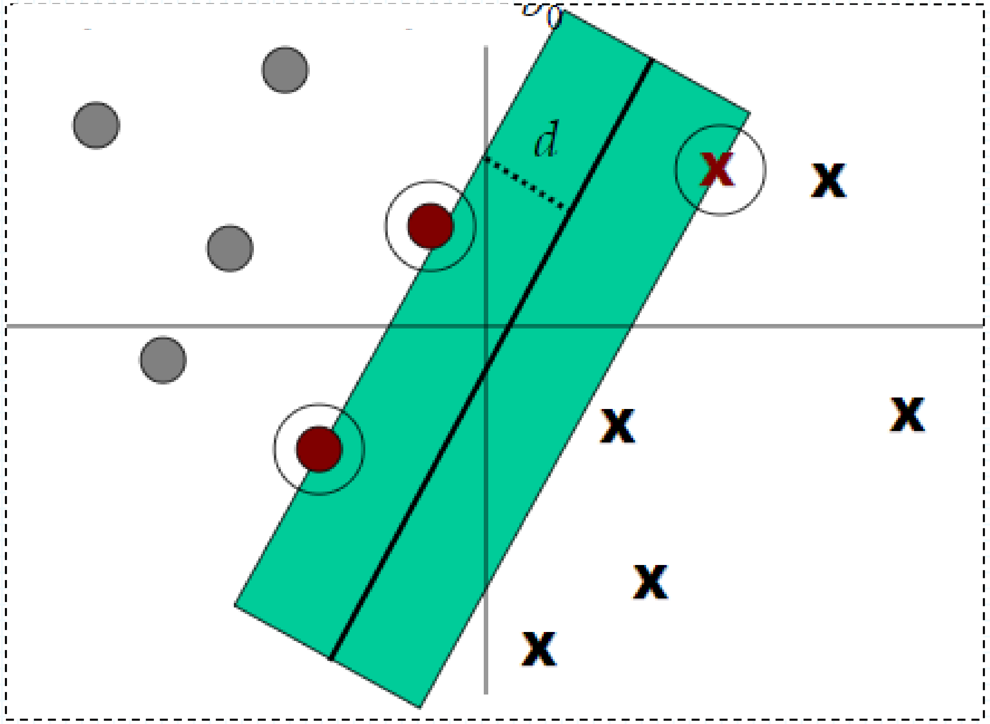
ML Pipeline


New input
Data
Feature Engineering
Training Approach
model
النموذج المتعلَّم
مجموعة معطيات
هندسة السمات
منهجية التعلم
Predicted output
الخرج المتوقع
مجموعة معطيات الاختبار
مجموعة معطيات التعلم

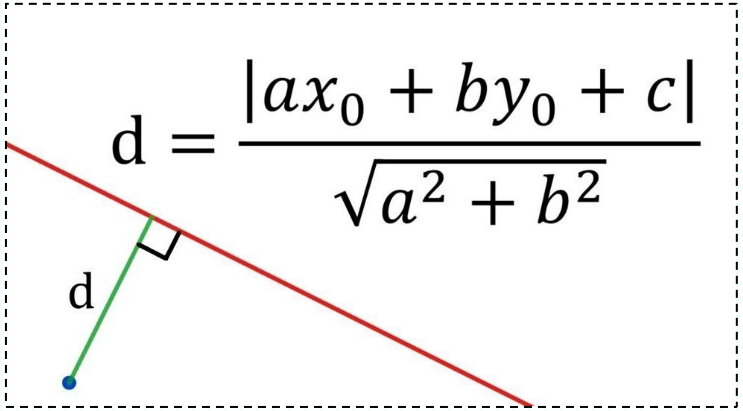
Intuition

- **Our problem:**

Maximizing the shortest distance to the closest positive or negative point.



The distance between a point (X_0,Y_0) and a line aX+bY+c=0



So.. What is our optimization problem?

- **Our problem:**

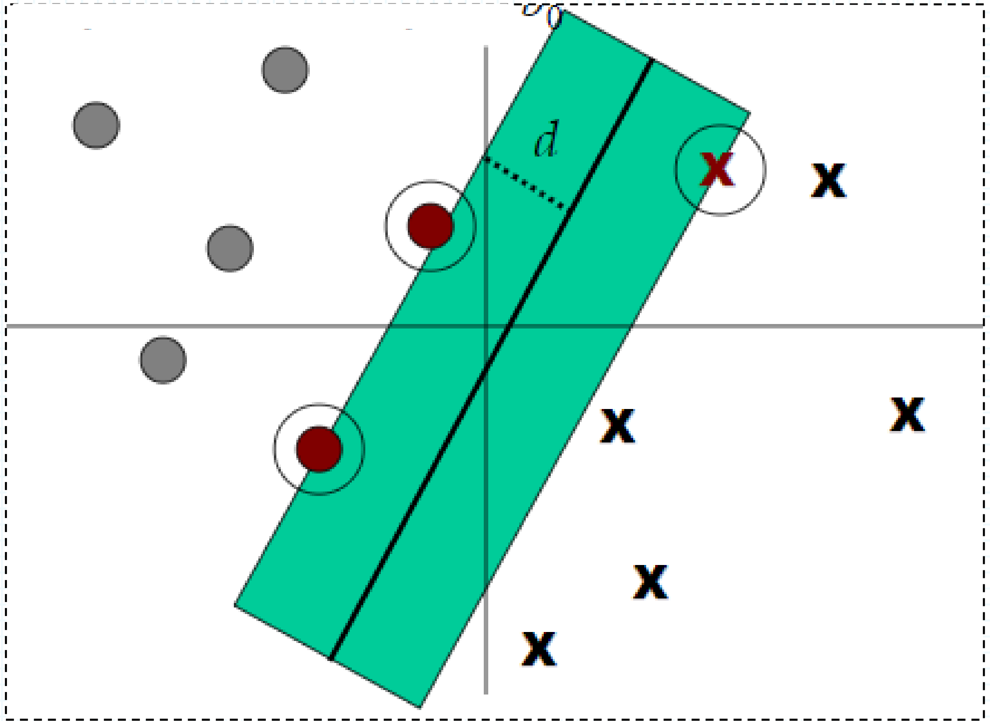
Maximizing the shortest distance to the closest positive or negative point.

In our case

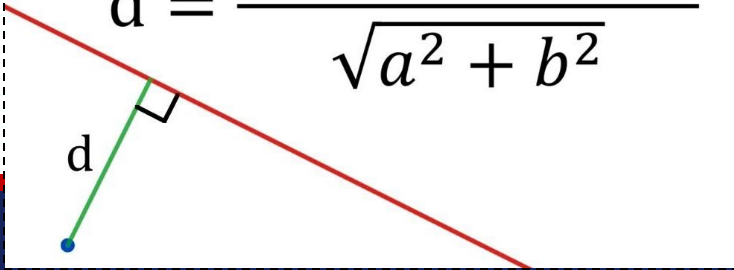
the distance of a hyperplane equation: $w^T\Phi(x) + b = 0$ from a given point vector $\Phi(x_0)$ can be easily written as :

$$d_H(\phi(x_0)) = \frac{|w^T(\phi(x_0)) + b|}{||w||_2}$$

$$||w||_2 =: \sqrt{w_1^2 + w_2^2 + w_3^2 + \dots w_n^2}$$



$$\mathrm{d} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

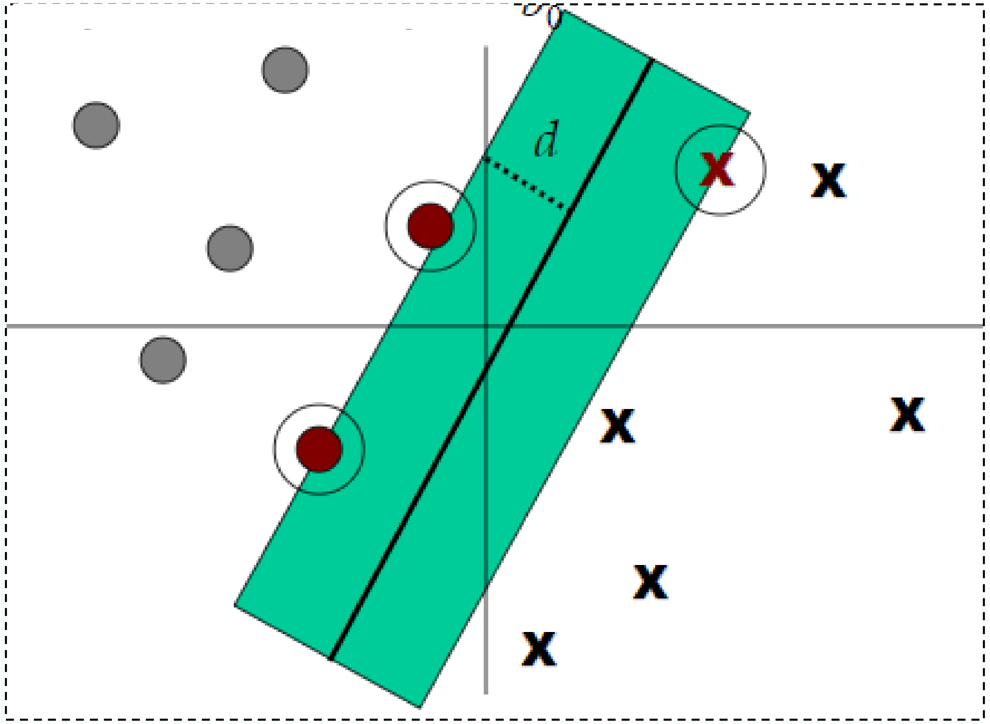


So.. What is our optimization problem?

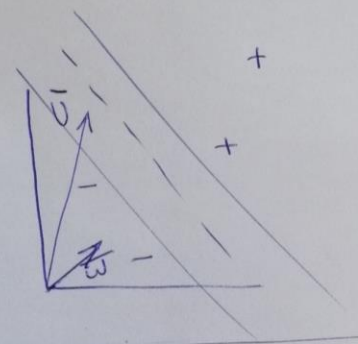
- **Our problem:**

Maximizing the shortest distance to the closest positive or negative point.

$$w^* = arg_w max\, [min_n\, d_H(\phi(x_n))]$$



Note that W represents all parameters i.e. w and b



$\bar{w}\cdot\bar{U} \geq C$ the N^+

$\bar{w}\cdot\bar{U} + b \geq \phi$

$\bar{w}\, x_+^* + b \geq 1$

$w\, x_- + b \leq -1$

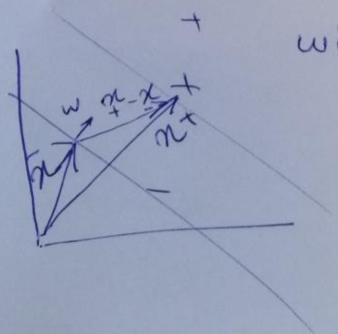
$y_i = +1$ for + samples

$\quad\; -1$ for − samples

constrait

$y_i(\bar{w}\cdot\bar{x}_i + b) \geq 1$ for all points

(*) $y_i(\bar{w}\,\bar{x}_i + b) = 1$ for support vector



width $= (\bar{x}_+ - \bar{x}_-)\cdot\dfrac{\bar{w}}{\|w\|}$

support vector

$= (\bar{x}_+\bar{w} - \bar{x}_-\bar{w})\dfrac{1}{\|w\|}$ (*)

from (*)

$x_+ \rightarrow \bar{w}x_+ + b = 1 \rightarrow \bar{w}x_+ = 1 - b$

$x_- \rightarrow \bar{w}x_- + b = -1 \rightarrow -\bar{w}x_- = 1 + b$

$\Downarrow$

width $= \dfrac{2}{\|w\|}$

Goal Maximize width $\dfrac{2}{\|w\|}$

$\Downarrow$

Minimize $\|w\|$

$\Downarrow$

Goal $\boxed{\text{Minimize } \frac{1}{2}\|w\|^2}$

true only if the constrait is satisfied

$\Downarrow$

use lagrange Multiplier

$L = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i\left[y_i(\bar{w}\,\bar{x}_i + b) - 1\right]$

Simplified Objective

$$\boxed{w^*, b^* = \arg\underset{w,b}{Min}\frac{1}{2}\|w\|^2\ ,\quad s.t.\ \ y_n\left(w^T\left(\emptyset(x_n)\right) + b\right) \geq 1 \quad \forall n}$$

How to find the optimum w and b?

SVM Optimization

$$w^*, b^* = \arg\underset{w,b}{Min} \frac{1}{2}\|w\|^2 , \quad s.t. \;\; y_n\left(w^T\big(\emptyset(x_n)\big) + b\right) \geq 1 \quad \forall n$$

Solved by Lagrange multiplier method:

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 \; - \sum_n \alpha_n [y_n\big(w^T\big(\emptyset(x_n)\big) + b\big) - 1]$$

where $\boldsymbol{\alpha}$ is the Lagrange multiplier

The optimization problem can be solved by setting **derivatives** of *Lagrangian* to 0

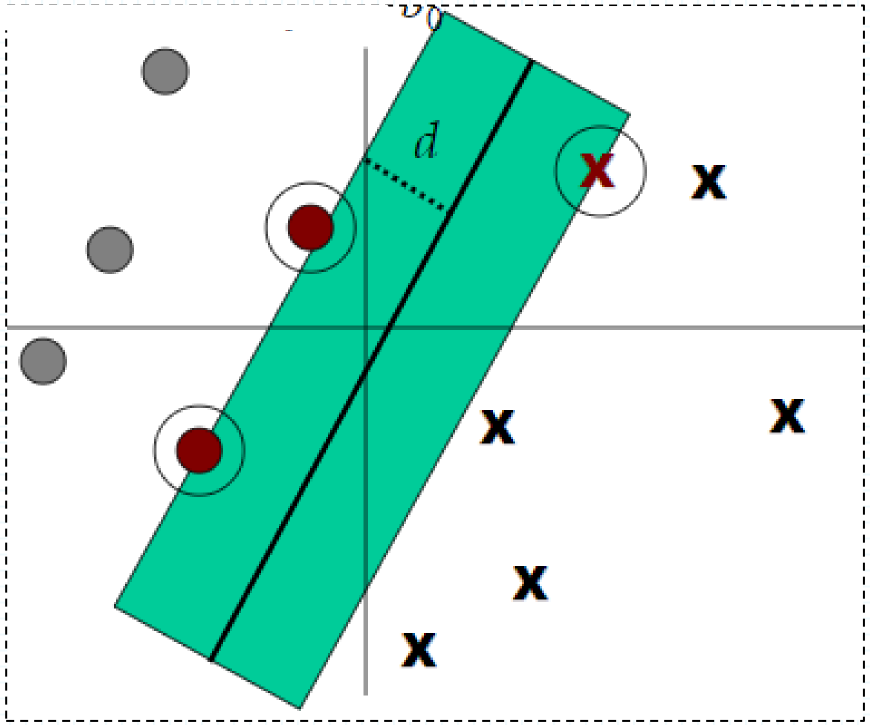
SVM Optimization

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 \; - \sum_{n} \alpha_n [y_n\big(w^T\big(\emptyset(x_n)\big) + b\big) - 1]$$

$$\begin{cases} \dfrac{\partial L}{\partial w} = w \; - \displaystyle\sum_{n} \alpha_n y_n \emptyset(x_n) \quad = 0 \;\; \Rightarrow \;\; w = \displaystyle\sum_{n} \alpha_n y_n \emptyset(x_n) \\ \dfrac{\partial L}{\partial b} = \displaystyle\sum_{n} \alpha_n y_n = 0 \;\; \Rightarrow \displaystyle\sum_{n} \alpha_n y_n \; = 0 \end{cases}$$

SVM Optimization

$$L = \sum_{n} \alpha_n + \frac{1}{2} \sum_{n,m} \alpha_n \alpha_m y_n y_m \emptyset^T(x_m) \emptyset(x_n)$$



The Boundary:

$$Y = w^T\big(\emptyset(x)\big) + b = \sum_n \alpha_n y_n \emptyset^T(x_n)\,\emptyset(x)$$

Only depend on inner products

SVM Optimization

$$w^*, b^* = \arg\underset{w,b}{Min}\frac{1}{2}\|w\|^2, \quad s.t.\ \ y_n(w^T(\emptyset(x_n)) + b) \geq 1 \quad \forall n$$

$$Y = w^T(\emptyset(x)) + b = \sum_n \alpha_n y_n \emptyset^T(x_n)\,\emptyset(x)$$

The decision rule in SVMs only depends on the dot product with support vectors

several important implications

Computational efficiency
Memory efficiency
Robustness to noise and outliers