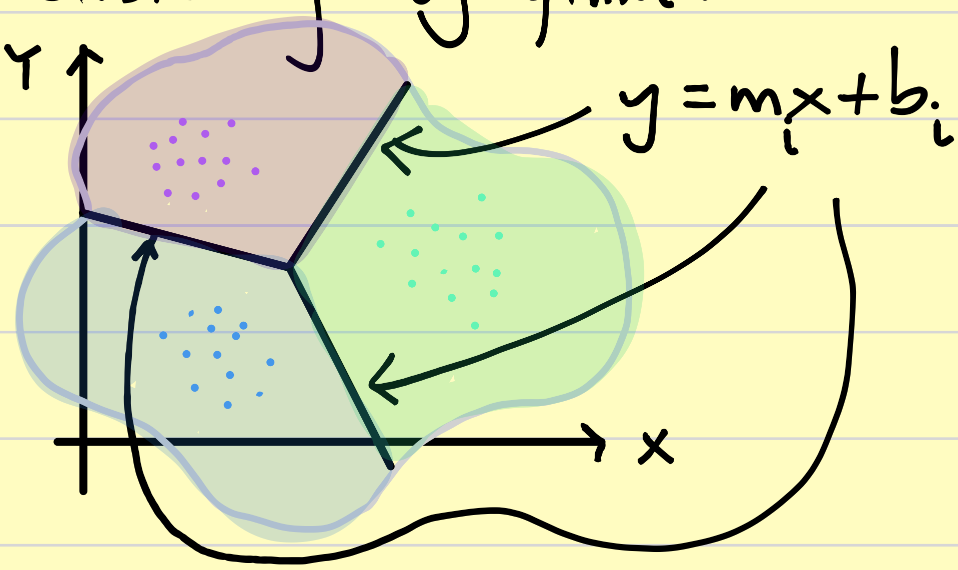
20241212_Business_Intelligence_WIN6

MACHINE LEARNING (ML)

SUPPORT VECTOR MACHINES (SVM)

- The goal of SVM ist to separate the space in regions. Based on the training dataset, we find with SVM the optima separating Line (2D), or separating plane (3D), or the separating hyperplane (more than 3D) that separates the space in regions.
- SVM is therefore a clustering algorythm.

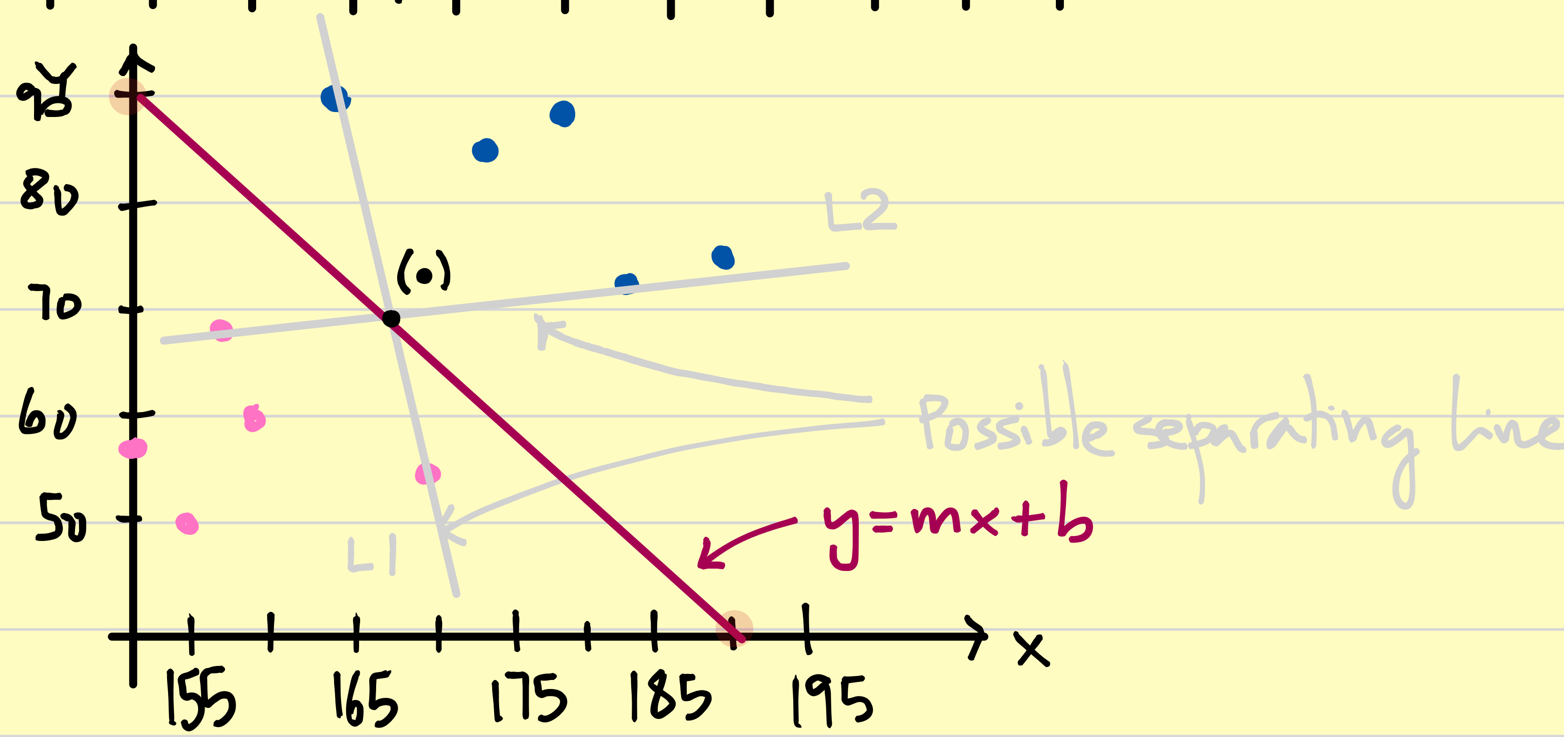
Example:



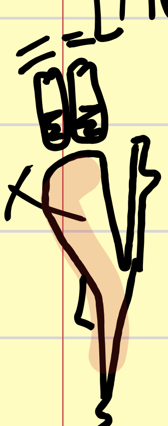
- We need training data to perform SVM. This means we talk about a „supervised" learning algorythm.
- We find the optimal separating hyperplane because we find the line with maximum distance to the training points.

Example.

Height	x	155	160	158	150	170	165	175	180	190	185
Weight	y	50	60	68	58	55	90	85	88	75	72
Sex		♀	♀	♀	♀	♀	♂	♂	♂	♂	♂

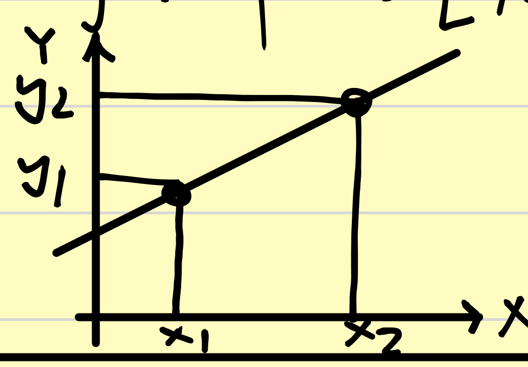


The best separating line ($y = mx + b$) goes through the points [190,0] and [0,90].



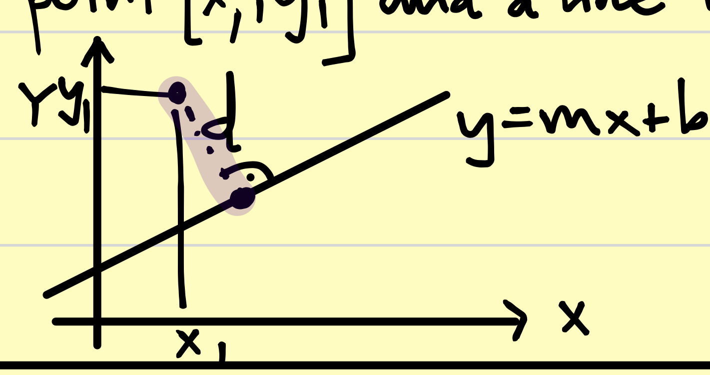
Equation of a line going through two points $[x_1, y_1]$ and $[x_2, y_2]$

$$\boxed{\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}}$$



The distance between a point $[x_1, y_1]$ and a line $y = mx + b$

$$\boxed{d = \frac{|mx_1 + y_1 + b|}{\sqrt{m^2+1}}}$$



Example: Given two data classes (red & blue), please perform all necessary calculations to:

a) find the optimal separating line ($y = mx + b$) to separate the two classes.

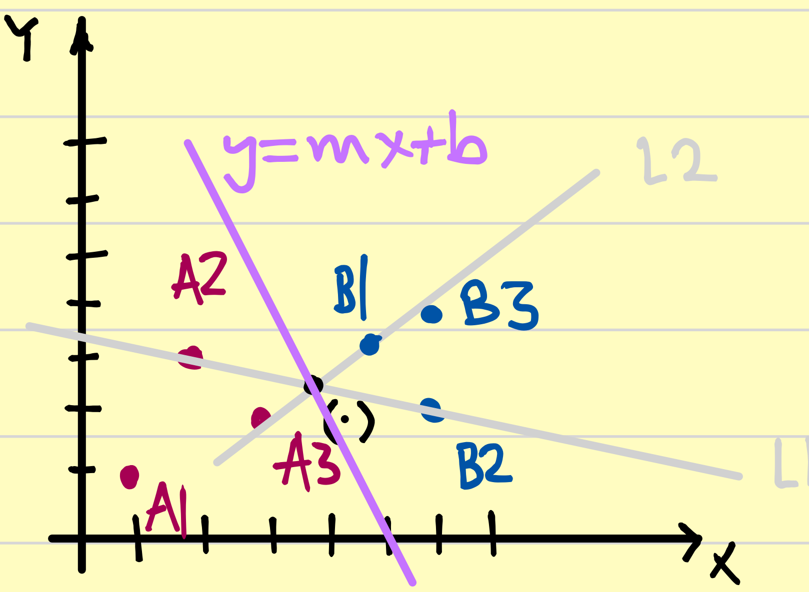
b) calculate the distance between the points closest to the line.

c) Identify the "support vectors".

RED (Class A): A1 [1,1], A2 [2,3], A3 [3,2]

BLUE (Class B): B1 [5,3], B2 [6,2], B3 [6,4]

step 1. Draw the data



step 2. calculate the point (•) where L1 & L2 intersect.

L1: A2 [2,3] B2 [6,2]: $\frac{y-3}{x-2} = \frac{2-3}{6-2} \rightarrow y = 3 - \frac{1}{4}(x-2)$ (1)

L2: A3 [3,2] B1 [5,3]: $\frac{y-2}{x-3} = \frac{3-2}{5-3} \rightarrow y = 2 + \frac{1}{2}(x-3)$ (2)

$$\boxed{\frac{y-2}{x-3} = \frac{1}{2} \longrightarrow y-2 = \frac{1}{2}(x-3) \longrightarrow y = 2 + \frac{1}{2}(x-3)}$$

$(1)-(2): \quad 0 = 3 - \frac{1}{4}(x_0 - 2) - 2 - \frac{1}{2}(x_0 - 3) \rightarrow$

$\rightarrow 0 = 1 - \frac{1}{4}x_0 + \frac{1}{2} - 2 - \frac{1}{2}x_0 + \frac{3}{2} \longrightarrow x_0 = 4$

$\underset{(1)}{\rightarrow} y_0 = 3 - \frac{1}{4}(4-2) = 2'5$

$(\circ)\ [x_0, y_0] = [4, 2'5] \longrightarrow 2'5 = m.4 + b \qquad (3)$

Step 3. calculate the distance btw the separating line and the two closest points. A3 & B1

$d_{L.A3} = \frac{|m.3 + 2 + b|}{\sqrt{m^2+1}}$ [3,2]

$d_{L.B1} = \frac{|m.5 + 3 + b|}{\sqrt{m^2+1}}$ [5,3]

Both distances should be equal to ensure max. distance

(4)

$m.3 + 2 + \cancel{b} = m.5 + 3 + \cancel{b}$

$\downarrow$

$m = -0'5$

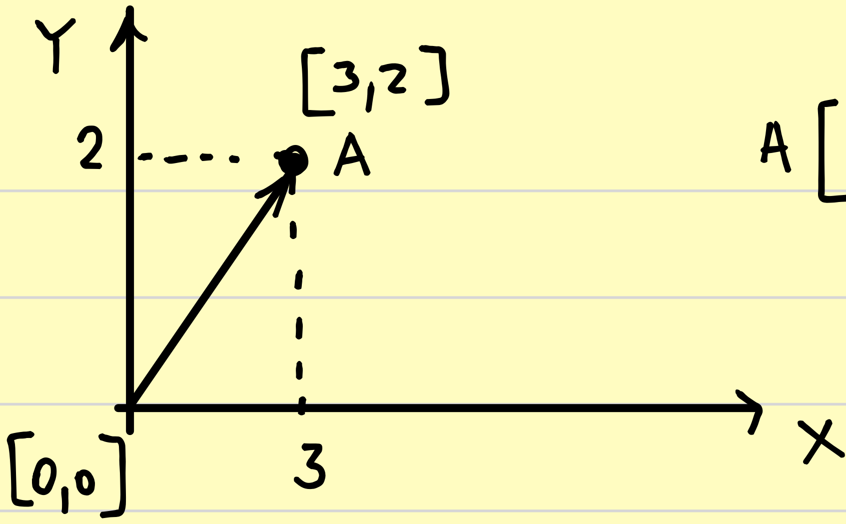
$\downarrow (3)$

$b = 4'5$

- Separating line: $\boxed{y = -0'5x + 4'5}$ ✓

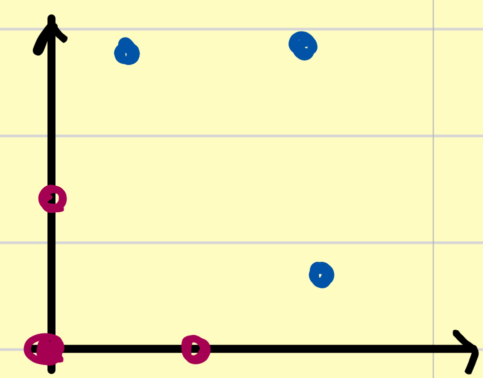
- $d_{L.A3} = d_{L.B1} = \frac{|3.(-0'5) + 2 + 4'5|}{\sqrt{(-0'5)^2 + 1}} = \frac{1}{\sqrt{\frac{5}{4}}} = \frac{5}{\sqrt{5}} = \frac{5\sqrt{5}}{5} = \sqrt{5}$ ✓

- Support vectors: A3[3,2] B1[5,3]



$$A[3,2] \equiv A[3-0, 2-0]$$

Example: Questions as above.



RED(A): $A_1[0,0]$ $A2[2,0]$ $A3[0,2]$

BLUE(B): $B_1[1,4]$ $B2[4,1]$ $B3[4,4]$

$$L1: \frac{y-1}{x-4} = \frac{2-1}{0-4} \rightarrow y = 1 - \frac{1}{4}(x-4) \qquad (1)$$

$$(-)$$

$$L2: \frac{y-0}{x-2} = \frac{4-0}{1-2} \rightarrow y = 0 - 4(x-2) \qquad (2)$$

$$0 = 1 - \frac{1}{4}(x_0 - 4) + 4(x_0 - 2) \rightarrow$$

$$\rightarrow 0 = 1 - \frac{1}{4}x_0 + 1 + 4x_0 - 8 \rightarrow x_0 = 1'6 \xrightarrow{(1)} y_0 = 1'6$$

$$(\bullet) \equiv [x_0, y_0] = [1'6, 1'6] \rightarrow 1'6 = m.1'6 + b \qquad (3)$$

$$d_{L.A3} = \frac{|m.0+2+b|}{\sqrt{m^2+1}}$$

$$d_{L.B2} = \frac{|m.4+1+b|}{\sqrt{m^2+1}}$$

$$\Rightarrow m.0+2+\cancel{b} = m.4+1+\cancel{b} \rightarrow m = 0'25$$

$$\downarrow (3)$$

$$b = 1'2$$

$$y = 0'25x + b$$

$$d_{L.A3} = \frac{|2+1'2|}{\sqrt{(0'25^2)+1}} = 3'0117$$

SUPPORT VECTORS: [A3] & [B2]