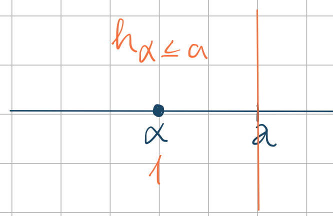
Task 1.

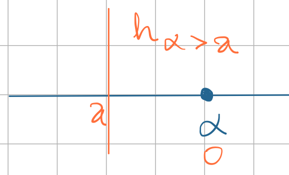
By definition, $\mathcal{H}_C = \{(h(c_1), \dots, h(c_m)) : h \in \mathcal{H}\}$, where $C = \{c_1, \dots, c_m\} \subset \mathcal{X}$ and $\mathcal{H}$ is a class of functions from $\mathcal{X}$ to $\{0, 1\}$.

Now, for $C_1 = \{\alpha\}$: $\quad \mathcal{H}_{C_1} = \{h_{\alpha \leq a}, h_{\alpha > a}\} = \{(1), (0)\}$

$|\mathcal{H}_{C_1}| = 2^{|C_1|} = 2^1 = 2 \rightarrow \mathcal{H}_{C_1}$ shatters C_1



set $a = \alpha + 1$ (i.e., $\alpha > a$) then $h_a(\alpha) = 1$

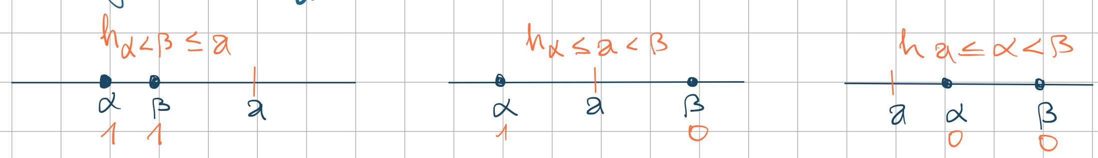


set $a = \alpha - 1$ (i.e, $\alpha < a$) then $h_a(\alpha) = 0$

- For $C_2 = \{\alpha, \beta\}$

set $a := \beta + 1 \Rightarrow \alpha < \beta \leq a \qquad h_{a=\beta+1}(\alpha, \beta) = (1, 1)$
set $a := \alpha - 1 \Rightarrow a < \alpha < \beta \qquad h_{a=\alpha-1}(\alpha, \beta) = (0, 0)$
set $\alpha \leq a < \beta \Rightarrow \qquad h_a(\alpha, \beta) = (1, 0)$

Therefore, $\mathcal{H}_{C_2} = \{(1,1), (0,0), (1,0)\}$.

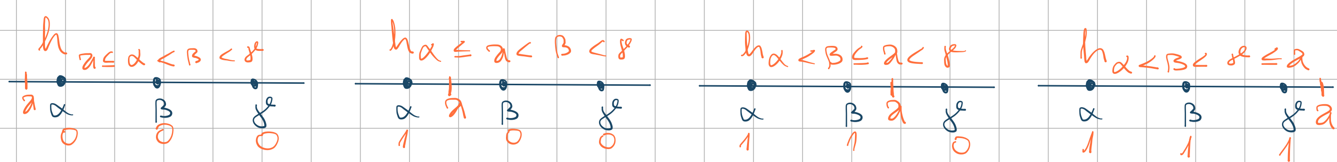


However, no $h \in \mathcal{H}$ can account for the labeling $(0, 1)$ because of the condition $\alpha < \beta$. Therefore, $\mathcal{H}$ does NOT shatter C_2.

Formally, $|\mathcal{H}_{C_2}| = 3 \neq 4 = 2^{|C_2|}$

- $C_3 = \{\alpha, \beta, \gamma\}$

$\mathcal{H}_{C_3} = \{(0,0,0), (1,0,0), (1,1,0), (1,1,1)\}$



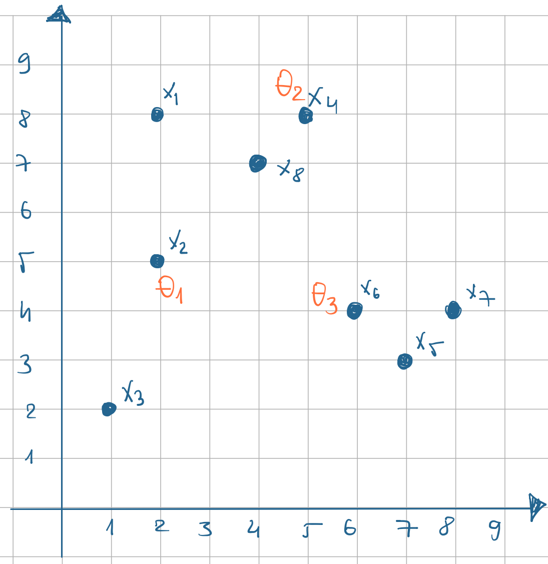
No $h \in \mathcal{H}$ can achieve the labellings $(0,1,0), (1,0,1), (0,0,1)$ and $(0,1,1)$. Therefore, $\mathcal{H}$ does NOT shatter C_3.

$|\mathcal{H}_{C_3}| = 4 \neq 8 = 2^{|C_3|}$

$\Rightarrow$ VC-dimensionality: $\quad \mathrm{VCdim}(\mathcal{H}) = 1$, because $C = \{\alpha\}$ is the maximal size of a set $C \subset \mathcal{X}$ that can be shattered by $\mathcal{H}$.

Task 2(a) –
$K=3$; Forgy initialization

~ ITERATION 1 ~



Random Initialization of clusters:

$$\theta_1 = x_2 = (2,5)$$
$$\theta_2 = x_4 = (5,8)$$
$$\theta_3 = x_6 = (6,4)$$

Assign data points based on distance:

- $x_1 = (2,8)$:

$$\|\theta_1 - x_1\| = \sqrt{(2-2)^2 + (5-8)^2} = \sqrt{9} = 3$$
$$\|\theta_2 - x_1\| = \sqrt{(5-2)^2 + (8-8)^2} = \sqrt{9} = 3$$
$$\|\theta_3 - x_1\| = \sqrt{(6-2)^2 + (4-8)^2} = \sqrt{16+16} = \sqrt{32} = 4\sqrt{2}$$

  - Since x_1 is equally close to θ_1 and θ_2, we assign it randomly to the cluster $\mathcal{M}_1$ with centroid θ_1. So far,

$$\mathcal{M}_1 = \{\underbrace{(2,5)}_{\text{centroid}}, (2,8)\}$$

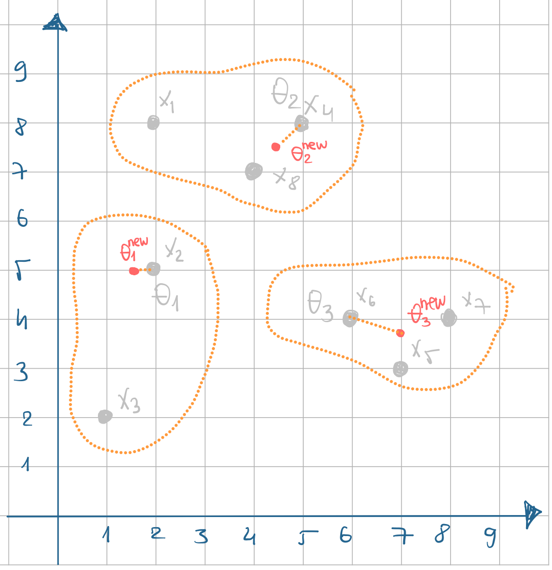
- $x_3 = (1,2)$:

$$\|\theta_1 - x_3\| = \sqrt{(2-1)^2 + (5-2)^2} = \sqrt{1+9} = \sqrt{10} \rightarrow \text{assign to } \mathcal{M}_1$$
$$\|\theta_2 - x_3\| = \sqrt{(5-1)^2 + (8-2)^2} = \sqrt{16+36} = \sqrt{52}$$
$$\|\theta_3 - x_3\| = \sqrt{(6-1)^2 + (4-2)^2} = \sqrt{25+4} = \sqrt{29}$$

$$\mathcal{M}_1 = \{(2,5), (2,8), (1,2)\}$$

- Similarly, we assign x_5 & x_7 to $\mathcal{M}_3$ and x_8 to $\mathcal{M}_2$.
  After the first iteration we have:

$$\mathcal{M}_1 = \{(2,5), (2,8), (1,2)\}$$
$$\mathcal{M}_2 = \{(5,8), (4,7)\}$$
$$\mathcal{M}_3 = \{(6,4), (7,3), (8,4)\}$$



Calculate new centroid coordinates:

- $\theta_1 = \left(\dfrac{2+2+1}{3}, \dfrac{5+8+2}{3}\right) = (1.67, 5)$
- $\theta_2 = \left(\dfrac{5+4}{2}, \dfrac{8+7}{2}\right) = (4.5, 7.5)$
- $\theta_3 = \left(\dfrac{6+7+8}{3}, \dfrac{4+3+4}{3}\right) = (7, 3.67)$

~ ITERATION 2 ~

Calculate the distance of each data point to the new centroids:

- x_1: $\|\theta_1 - x_1\|_2 = \sqrt{0{,}11+9} = \sqrt{9{,}11} = 3{,}018$
  $\|\theta_2 - x_1\|_2 = \sqrt{6.25+0{,}25} = \sqrt{6{,}5} = 2{,}54$ → move x_1 from $\mathcal{M}_1$ to $\mathcal{M}_2$
  $\|\theta_3 - x_1\|_2 = \sqrt{25+18{,}75} = \sqrt{43{,}74} = 6{,}61$

Now, $\mathcal{M}_1 = \{(2,5),(1,2)\}$
$\mathcal{M}_2 = \{(2,8),(5,8),(4,7)\}$
$\mathcal{M}_3 = \{(7,3),(6,4),(8,4)\}$

Calculate the new centroids:

$$\theta_1 = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = \left(\frac{3}{2}, \frac{7}{2}\right) = (1.5, 3.5)$$

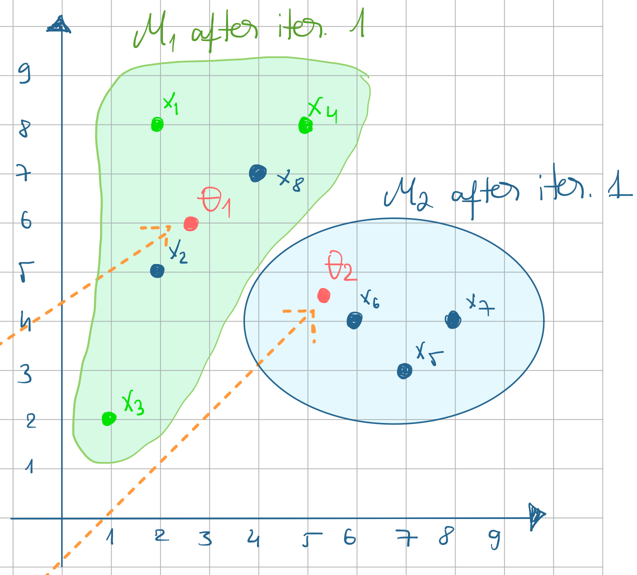
$$\theta_2 = \left(\frac{2+5+4}{3}, \frac{8+8+7}{3}\right) = \left(\frac{11}{3}, \frac{23}{3}\right) = (3.67, 7.67)$$

$$\theta_3 = \left(\frac{7+6+8}{3}, \frac{3+4+4}{3}\right) = \left(7, \frac{11}{3}\right) = (7, 3.67)$$

Task 2(b)
K=2; random partition method

Initial partition: (0, 1, 0, 1, 0, 1, 1, 1)

$\mathcal{M}_1 = \{x_1, x_3, x_4\}$
$\mathcal{M}_2 = \{x_2, x_4, x_5, x_6, x_7\}$



~ ITERATION 1 ~

$$\theta_1 = \left(\frac{2+1+5}{3}, \frac{8+2+8}{3}\right) = (2.67, 6)$$

$$\theta_2 = \left(\frac{2+4+6+7+8}{5}, \frac{5+7+4+3+4}{5}\right) = (5.4, 4.6)$$

Calculate distances from each x_i to θ_1 & θ_2

$\|\theta_1 - x_3\| = \sqrt{(2.67-1)^2+(6-2)^2} = \sqrt{2.79+16} = 4{,}33$ → assign x_3 to $\mathcal{M}_1$

$\|\theta_2 - x_3\| = \sqrt{(5.4-1)^2+(4.6-2)^2} = \sqrt{19{,}36+6{,}76} = 5{,}11$

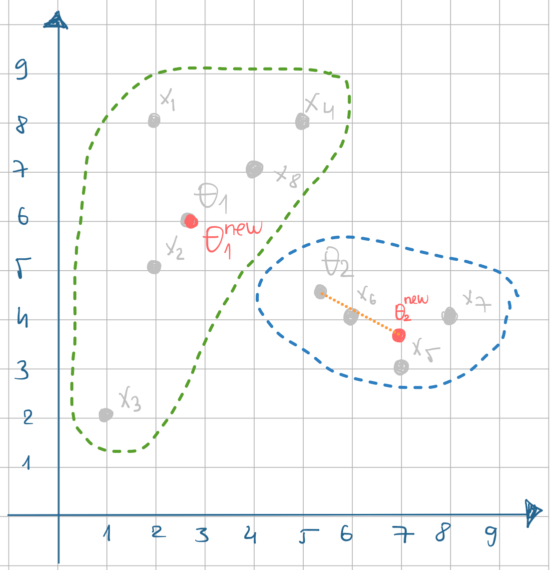
- After performing the calculations for each x_i:

$$\mathcal{M}_1 = \{x_1, x_2, x_3, x_4, x_8\}$$
$$\mathcal{M}_2 = \{x_5, x_6, x_7\}.$$

~ ITERATION 2 ~

$$\theta_1 = \left(\frac{2+2+1+5+4}{5}, \frac{8+5+2+8+7}{5}\right) = (2.8,\ 6)$$

$$\theta_2 = \left(\frac{7+6+8}{3}, \frac{3+4+4}{3}\right) = (7,\ 3.67)$$



Calculating distances of x_i from θ_1, θ_2:

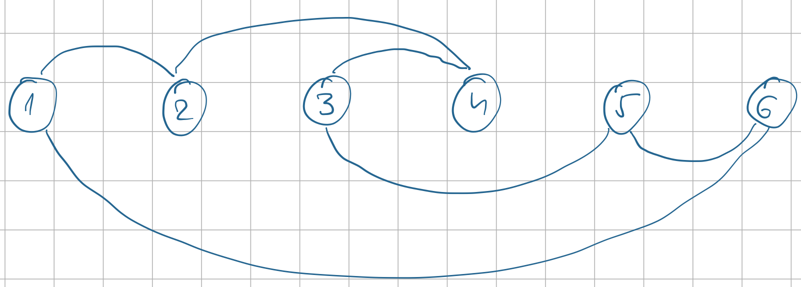
- The new centroids don't change any of the assignments:

$$\mathcal{M}_1 = \{x_1, x_2, x_3, x_4, x_8\}$$

$$\mathcal{M}_2 = \{x_5, x_6, x_7\}$$

Task 3: mutual KNN graph for k=2

0,9	1 − 2
0,9	5 − 6
0,8	2 − 4
0,7	3 − 5
0,6	3 − 4
0,5	1 − 3
0,4	2 − 3
0,4	4 − 5
0,3	4 − 6
0,2	1 − 6



KNN graph for k=2