Ex 4.

Round 1: Center is undeslined

a) Cluster 1: <u>A</u>

Clustes 2: <u>B</u>

Clustes 3: <u>C</u>, D, E, F, G, H

Round 2:

Clustes 1: <u>A</u>

Clustes 2: <u>B</u>, C

Clustes 3: <u>(7.5, 6)</u>, D, E, F, G, H

Round 3:

Clustes 1: <u>A</u>, B

Clustes 2: <u>(3, 9.5)</u>, C

Cluster 3: <u>(8.5, 5.6)</u>, D, E, F, G, H

Round 4:

Clustes 1: <u>(2.5, 11.5)</u>, A, B

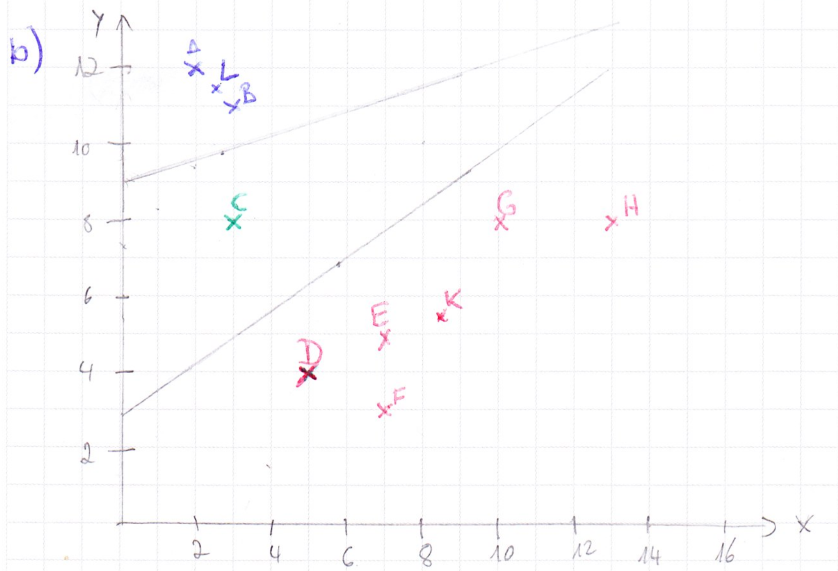
Clustes 2: <u>C</u>

Clustes 3: <u>(8.5, 5.6)</u>, D, E, F, G, H

Stays the same as in Round 3

↳ Alg. determines

b)



L is mean of clustes 1: (2.5, 11.5)

C is mean of clustes 2: (3, 8)

K is mean of clustes 3: (8.5, 5.6)

b) $\| (x, f(x)) - L \| = \| (x, f(x)) - C \|$

$\Leftrightarrow \| (x - L_1, f(x) - L_2) \| = \| (x - C_1, f(x) - C_2) \|$

$\Leftrightarrow \sqrt{(x-L_1)^2 + (f(x) - L_2)^2} = \sqrt{(x - C_1)^2 + (f(x) - C_2)^2}$

$\Leftrightarrow (x - L_1)^2 + (f(x) - L_2)^2 = (x - C_1)^2 + (f(x) - C_2)^2$

$\Leftrightarrow (f(x) - L_2)^2 - (f(x) - C_2)^2 = (x - C_1)^2 - (x - L_1)^2$

$\Leftrightarrow -2f(x)L_2 + 2f(x)C_2 + L_2^2 + C_2^2 = (x - C_1)^2 - (x - L_1)^2$

$\Leftrightarrow 2(f(x)C_2 - f(x)L_2) = (x - C_1)^2 - (x - L_1)^2 - L_2^2 C_2^2$

$\Leftrightarrow f(x) = \dfrac{(x - C_1)^2 - (x - L_1)^2 - L_2^2 - C_2^2}{2(C_2 - L_2)}$

$\Leftrightarrow f(x) = \dfrac{-2C_1 x + 2L_1 x + C_1^2 + L_1^2 - L_2^2 - C_2^2}{2(C_2 - L_2)}$ this is the general formula

This leads to the 3 functions:

Between Cluster 1 and Cluster 2: $f(x) = \frac{1}{7}x + \frac{181}{7}$

Between Cluster 2 and Cluster 3: $f(x) = \frac{11}{4{,}8}x + \frac{14{,}11}{4{,}8}$

Between Cluster 1 and Cluster 3: $f(x) = \frac{12}{11{,}8}x + \frac{85{,}11}{11{,}8}$

c) Choose A, F and H as initial cluster means. Underlined is the center

$\|B-A\| = \sqrt{2} \rightarrow A$ $\|C-A\| = \sqrt{17} \rightarrow A$ $\|D-A\| = \sqrt{73}$ $\|E-A\| = \sqrt{74}$ $\|G-A\| = 4\sqrt{5}$

$\|B-F\| = 4\sqrt{5}$ $\|C-F\| = \sqrt{41}$ $\|D-F\| = \sqrt{5} \rightarrow F$ $\|E-F\| = 2 \rightarrow F$ $\|G-F\| = \sqrt{34}$

$\|B-H\| = \cancel{2\sqrt{3}}\ \sqrt{103}$ $\|C-H\| = 10$ $\|D-H\| = 4\sqrt{5}$ $\|E-H\| = 3\sqrt{5}$ $\|G-H\| = 3 \rightarrow H$

Cluster 1: $\underline{A}$, B, C

Cluster 2: D, E, $\underline{F}$

Cluster 3: G, $\underline{H}$

With the formula $z^j \leftarrow \dfrac{\sum_{x \in C^j} x}{|C^j|}$ for all $j \in [k]$ the new centroids are:

For Cluster 1: (2,667, 10,333)

For Cluster 2: (6,333, 4)

For Cluster 3: (11,5, 8)

ALTRAN

c) With the computations with the new centroids it leads to the following clusters: The center is underlined.

Cluster 1: $\underline{(2{,}667, 10{,}333)}$, A, B, C

Cluster 2: $\underline{(6{,}333, 4)}$, D, E, F

Cluster 3: $\underline{(11{,}5, 8)}$, G, H

So the clusters are the same as in the round before. So the algorithm stopps

So we got a different set of final clusters as in a)

d) $\|A-G\| = 4\sqrt{5}$

$\|B-A\| = \sqrt{2}$

$\|B-G\| = \sqrt{58}$

The shortest distance of all the points to each other is the distance between A and B. So A and B has to be in a cluster except A and B are Centers of one cluster. But then the distance between A and G is longer then B and G. So G would be in the cluster of B.

So it is impossible that the k-mean Algorithm returns {A, G} as a cluster.