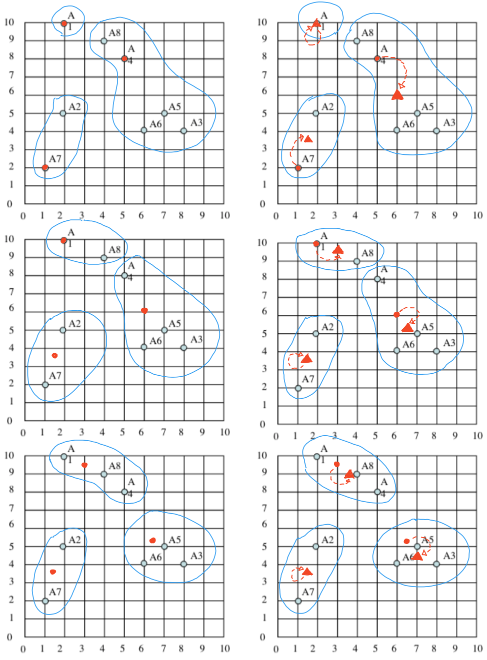
$d(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$

(a_1, a_2) (b_1, b_2)

https://miro.com/app/board/o9J_lEGehhs=/

Flat clustering (Chapter 16)

Exercise 16/1
Use the K-means algorithm with Euclidean distance to cluster the following N = 8 examples into K = 3 clusters:
A1 = (2, 10), A2 = (2, 5), A3 = (8, 4),
A4 = (5, 8), A5 = (7, 5), A6 = (6, 4),
A7 = (1, 2), A8 = (4, 9).
Suppose that the initial seeds (centers of each cluster) are: A1, A4 and A7.
Run the K-means algorithm for 3 epochs.
After each epoch, draw a 10 × 10 space with all the 8 points and show the clusters with the new centroids.



reassigning vectors   recomputing centroids

Algorithm 1 K-means($\{\vec{x}_1, \ldots, \vec{x}_N\}, K$, stopping criterion)

$(\vec{s}_1, \ldots, \vec{s}_K) \leftarrow$ SelectRandomSeeds($\{\vec{x}_1, \ldots, \vec{x}_N\}, K$)
for $k \leftarrow 1$ to K **do**
  $\vec{\mu}_k \leftarrow \vec{s}_k$
end for
repeat
  for $k \leftarrow 1$ to K **do**
    $\omega_k \leftarrow \{\}$
  end for
  for $n \leftarrow 1$ to N **do**
    $j \leftarrow \text{argmin}_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$
    $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$ ▷ reassigning vectors
  end for
  for $k \leftarrow 1$ to K **do**
    $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$ ▷ recomputing centroids
  end for
until a stopping criterion has been met
return $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}$

HIGH intra-cluster
LOW inter-cluster
similarity

Exercise 16/2
What makes a good clustering? Give some clustering evaluation metrics.