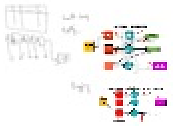
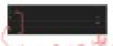


K means Clustering

- Definition
- Algo → [illegible]
- Steps

1. Randomize centroids



A (1,2) D (6,7) centroids

2. Assign each point to nearest centroid

$C_1(1,2) - A = 0$

$C_1 - B = 2$

$C_1 - C = 5.66$

$C_1 - D = 7.07$

$C_2(6,7) - A = 7.07$

$C_2 - B = 5.83$

$C_2 - C = 1.41$

$C_2 - D = 0$

C_1

AB

DC

C_2

3 New Centroids are calculated

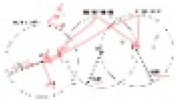
$C_1 \; A(1,2) B(1,4) \Rightarrow \left[\frac{1+1}{2}, \frac{2+4}{2}\right] (1, 3)$

$C_2 \; C(5,6) D(6,7) \Rightarrow \left[\frac{5+6}{2}, \frac{6+7}{2}\right] (5.5, 6.5)$

4 Repeat 2 to 3

DBSCAN

Density based spatial clustering of applications with noise



Hierarchical Clustering

Agglomerative clustering

1. Start with each data point as its own cluster
2. merge the two closest clusters based on a linkage criterion
3. Repeat merging until:
   All points form a single cluster or
   a stopping condition (number of clusters) is met.

Linkage criteria

Single Linkage: Minimum distance between points in the two clusters.

Complete Linkage: Maximum distance

Average Linkage: Average of all pairwise distances

Ward's Method: Minimizes variance between clusters