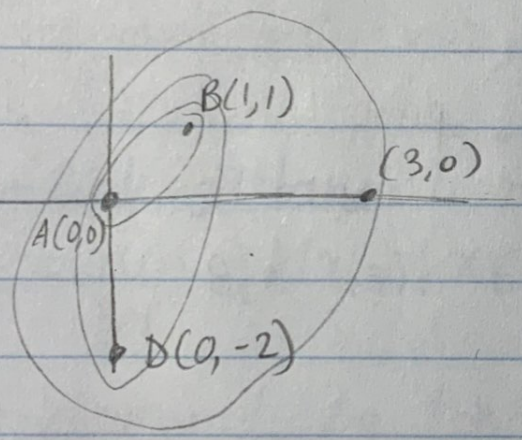


Hierarchical Clustering

P_1, P_2 points

$d(P_1, P_2)$ = euclidean distance

C_1, C_2 clusters

$D(C_1, C_2)$ = single-link (min distance)

Distance Matrix

	A	B	C	D
A	0	$\sqrt{2}$	3	2
B	$\sqrt{2}$	0	$\sqrt{5}$	$\sqrt{10}$
C	3	$\sqrt{5}$	0	$\sqrt{13}$
D	2	$\sqrt{10}$	$\sqrt{13}$	0

min d = d(A, B)

* merge A & B

⟶

	A&B	C	D
A&B	0	$\sqrt{5}$	2
C	$\sqrt{5}$	0	$\sqrt{13}$
D	2	$\sqrt{13}$	0

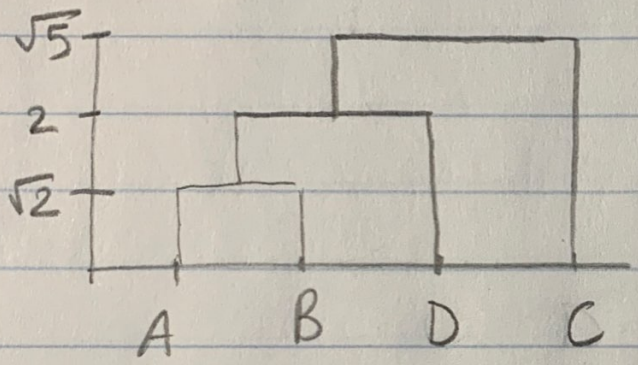
Scanned with CamScanner

min $d = d(A,B,D)$

*merge A, B and D

	A & B & D	C
A & B & D	0	$\sqrt{5}$
C	$\sqrt{5}$	0

Dendogram:



Density-Based Clustering:

- Density: if there are at least min-pts number of points in a ball of radius ε^*, then this region is dense *ε-neighborhood
- Core point: if there are at least min-pts in its ε-neighborhood
- Border point: is not a core point but is in the ε-neighborhood of a core point
- Noise points: neither core nor border

DBSCAN Algorithm

- label all points as core, border, or noise
- eliminate noise points
- put an edge between all core points that are within ε's of each other

Scanned with CamScanner