DBSCAN Algorithm (Density-based Spatial clustering of Applications with noise)

Example

P_1	4.5	8
P_2	5	7
P_3	6	6.5
P_4	7	5
P_5	9	4
P_6	7	3
P_7	8	3.5
P_8	9	5
P_9	4	4
P_{10}	3	7.5
P_{11}	4	6
P_{12}	3.5	5

→ Number of points within a specified radius

set $\epsilon = 1.9$, Minpts $= 4$

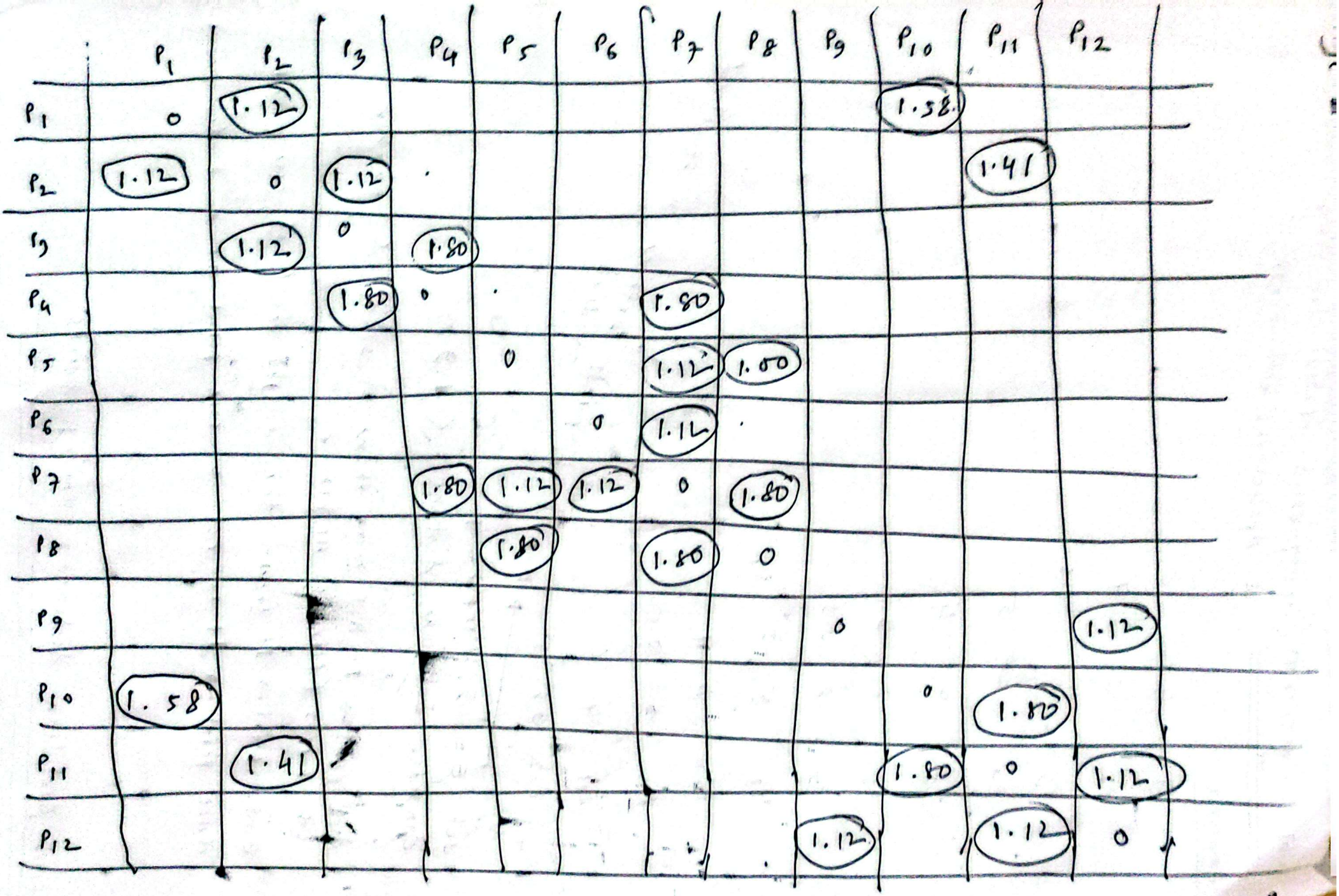
consider Eclidean distance ≤ 1.9

Core point → It has more than a specified number of points (Minpts) within ϵ

Border point → It is in the neighborhood of core point

Noise point → It is not a core point or a border point

Find Euclidean distances among points


P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
P1 0 1.12 1.58
P2 1.12 0 1.12 1.41
P3 1.12 0 1.80
P4 1.80 0 1.80
P5 0 1.12 1.00
P6 0 1.12
P7 1.80 1.12 1.12 0 1.80
P8 1.80 1.80 0
P9 0 1.12
P10 1.58 0 1.80
P11 1.41 1.80 0 1.12
P12 1.12 1.12 0

.ge the point

$P_1 \rightarrow P_2\ P_{10}$	$P_1\ P_2\ P_{10}$
$P_2 \rightarrow P_1\ P_3\ P_{11}$	$P_2\ P_1\ P_3\ P_{11}$
$P_3 \rightarrow P_2\ P_4$	$P_3\ P_2\ P_4$
$P_4 \rightarrow P_3\ P_7$	$P_4\ P_3\ P_7$
$P_5 \rightarrow P_7\ P_8$	$P_5\ P_7\ P_8$
$P_6 \rightarrow P_7$	$P_6\ P_7$
$P_7 \rightarrow P_4\ P_5\ P_6\ P_8$	$P_7\ P_4\ P_5\ P_6\ P_8$
$P_8 \rightarrow P_5\ P_7$	$P_8\ P_5\ P_7$
$P_9 \rightarrow P_{12}$	$P_9\ P_{12}$
$P_{10} \rightarrow P_1\ P_{11}$	$P_{10}\ P_1\ P_{11}$
$P_{11} \rightarrow P_2\ P_{10}\ P_{12}$	$P_{11}\ P_2\ P_{10}\ P_{12}$
$P_{12} \rightarrow P_9\ P_{11}$	$P_{12}\ P_9\ P_{11}$

Identifying core, [redacted] and Noise

$P_1\ P_2\ P_{10}$ → Noise

$P_2\ P_1\ P_3\ P_{11}$ ⇒ Core [since greater than or equal to 4]

$P_3\ P_2\ P_4$ → Noise

$P_4\ P_3\ P_7$ → Noise

$P_5\ P_7\ P_8$ → Noise

$P_6\ P_7$ → Noise

$P_7\ P_4\ P_5\ P_6\ P_8$ ⇒ core [since greater than or equal to 4]

$P_8\ P_5\ P_7$ → Noise

$P_9\ P_{12}$ → Noise

$P_{10}\ P_1\ P_{11}$ → Noise

$P_{11}\ P_2\ P_{10}\ P_{12}$ ⇒ Core [since greater than or equal to 4]

$P_{12}\ P_9\ P_{11}$ → Noise

P_1 P_2 P_{10}	P_1	Noise	Boorder — since P_1 is in P_2 P_1 P_3 P_{11}
P_2 P_1 P_3 P_{11}	P_2	core	core
P_3 P_2 P_4	P_3	Noise	Boorder — since P_3 is in P_2 P_1 P_3 P_{11}
P_4 P_3 P_7	P_4	Noise	Border — since P_4 is in P_7 P_4 P_5 P_6 P_8
P_5 P_7 P_8	P_5	Noise	Border — since P_5 is in P_7 P_4 P_5 P_6 P_8
P_6 P_7	P_6	Noise	Border — since P_6 is in P_7 P_4 P_5 P_6 P_8
P_7 P_4 P_5 P_6 P_8	P_7	core	~~Border~~ core
P_8 P_5 P_7	P_8	Noise	Border — since P_8 is in P_7 P_4 P_5 P_6 P_8
P_9 P_{12}	P_9	Noise	Noise
P_{10} P_1 P_{11}	P_{10}	Noise	Border — since P_{10} is in P_{11} P_2 P_{10} P_{12}
P_{11} P_2 P_{10} P_{12}	P_{11}	core	~~Border~~ core
P_{12} P_9 P_{11}	P_{12}	Noise	Border — since P_{12} is in P_{11} P_2 P_{10} P_{12}

conclusion P_9 is finally noise, so it is outlier data