Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Thomas Seidl
Janina Sontheim, Maximilian Hünemörder

Knowledge Discovery and Data Mining I
WS 2019/20

Exercise 9: Outlier Scores

Exercise 9-1   Monotonicity of Simple Outlier Scores

Proof or give an counterexample for the following claims:

(a) If o is an $D(\epsilon, \pi)$-outlier, it is also an $D(\epsilon', \pi)$-outlier for $\epsilon' \leq \epsilon$.

(b) If o is an $D(\epsilon, \pi)$-outlier, it is also an $D(\epsilon, \pi')$-outlier for $\pi' \geq \pi$.

(c) If o is an kNN-outlier for threshold τ, it is also an k'NN-outlier for the same threshold with $k' > k$.

(d) If o is an kNN-outlier for threshold τ, it is also an kNN-outlier for threshold $\tau' < \tau$.

(e) The local density is monotonously decreasing in k, i.e. $ld_k(o) \geq ld_{k'}(o)$ for $k' > k$.

Exercise 9-2   Outlier Scores

Given the following 2 dimensional data set:
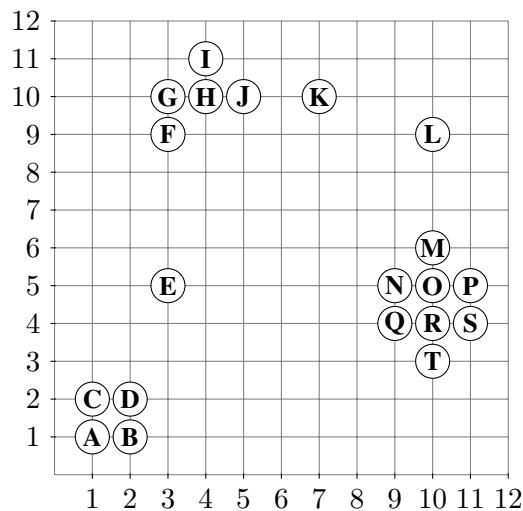
As distance function, use Manhattan distance $L_1(a, b) := |a_1 - b_1| + |a_2 - b_2|$. The following table summarises the pairwise distances.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	0	1	1	2	6	10	11	12	13	13	15	17	14	12	13	14	11	12	13	11
B	1	0	2	1	5	9	10	11	12	12	14	16	13	11	12	13	10	11	12	10
C	1	2	0	1	5	9	10	11	12	12	14	16	13	11	12	13	10	11	12	10
D	2	1	1	0	4	8	9	10	11	11	13	15	12	10	11	12	9	10	11	9
E	6	5	5	4	0	4	5	6	7	7	9	11	8	6	7	8	7	8	9	9
F	10	9	9	8	4	0	1	2	3	3	5	7	10	10	11	12	11	12	13	13
G	11	10	10	9	5	1	0	1	2	2	4	8	11	11	12	13	12	13	14	14
H	12	11	11	10	6	2	1	0	1	1	3	7	10	10	11	12	11	12	13	13
I	13	12	12	11	7	3	2	1	0	2	4	8	11	11	12	13	12	13	14	14
J	13	12	12	11	7	3	2	1	2	0	2	6	9	9	10	11	10	11	12	12
K	15	14	14	13	9	5	4	3	4	2	0	4	7	7	8	9	8	9	10	10
L	17	16	16	15	11	7	8	7	8	6	4	0	3	5	4	5	6	5	6	6
M	14	13	13	12	8	10	11	10	11	9	7	3	0	2	1	2	3	2	3	3
N	12	11	11	10	6	10	11	10	11	9	7	5	2	0	1	2	1	2	3	3
O	13	12	12	11	7	11	12	11	12	10	8	4	1	1	0	1	2	1	2	2
P	14	13	13	12	8	12	13	12	13	11	9	5	2	2	1	0	3	2	1	3
Q	11	10	10	9	7	11	12	11	12	10	8	6	3	1	2	3	0	1	2	2
R	12	11	11	10	8	12	13	12	13	11	9	5	2	2	1	2	1	0	1	1
S	13	12	12	11	9	13	14	13	14	12	10	6	3	3	2	1	2	1	0	2
T	11	10	10	9	9	13	14	13	14	12	10	6	3	3	2	3	2	1	2	0

(a) Calculate the $D(\epsilon, \pi)$-outliers using

   (i) $\epsilon = 2$ with $n\pi = 1$ and $n\pi = 2$.

  (ii) $\epsilon = 4$ with $n\pi = 1$, $n\pi = 3$ and $n\pi = 4$.

 (iii) $\epsilon = 6$ with $n\pi = 4$, $n\pi = 5$ and $n\pi = 6$.

(b) Calculate the kNN based outliers for $(k, \tau) = (3, 3)$ and $(k, \tau) = (5, 8)$. The point itself is counted as the 0-nearest neighbour.

(c) Given the following curves of the local density ld_k for different values of k.
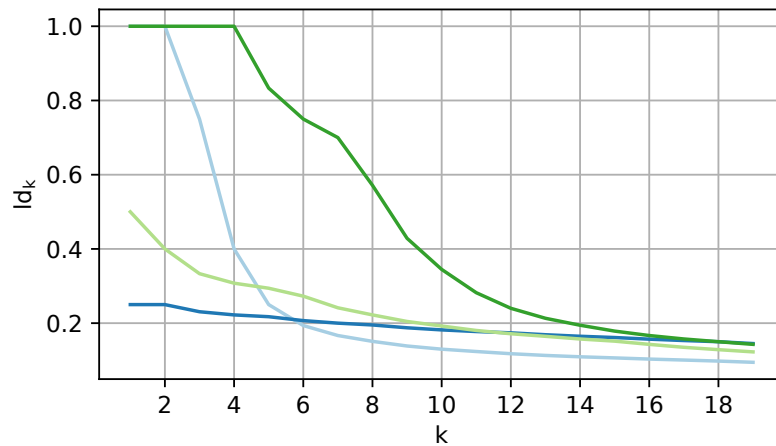
Can you identify which curve belongs to which point? Explain your mapping.