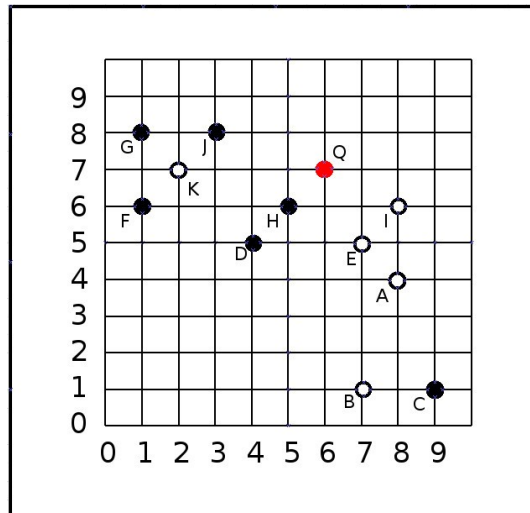


Κατηγοριοποίηση k εγγύτερων γειτόνων και Τεχνικές μείωσης του όγκου των δεδομένων (Data Reduction)

Στο σχήμα που σας δίνεται, φαίνονται στο Ευκλείδιο επίπεδο 11 σημεία που ανήκουν στις κλάσεις “μαύρος κύκλος” και “λευκός κύκλος”. Για ευκολία έχουμε δώσει ονόματα στα σημεία (γράμματα Α ως Κ). Μετρική απόστασης είναι η Ευκλείδια απόσταση. Για ευκολία τα 11 σημεία έχουν ακέραιες συντεταγμένες και στην πράξη δεν απαιτείται να υπολογίσετε τις αποστάσεις (ο εγγύτερος γείτονας φαίνεται με το “μάτι”). Το σημείο Q δεν ανήκει στο σύνολο δεδομένων εκπαίδευσης. Το σημείο Q πρέπει να κατηγοριοποιηθεί σε μια από τις δύο κλάσεις χρησιμοποιώντας τον κατηγοριοποιητή των k εγγύτερων γειτόνων.



Ερωτήσεις

1. Σε ποια κλάση θα κατηγοριοποιηθεί το Q αν ορίσουμε $\alpha. k=1$ και $\beta. k=3$
2. Εφαρμόστε τον αλγόριθμο editing ENN-rule με $k=3$ και δώστε το ES (edited set) που θα προκύψει
3. Σε ποια κλάση θα κατηγοριοποιηθεί το Q αν χρησιμοποιήσουμε το ES που προέκυψε από το ερώτημα 2 και ορίσουμε $\alpha. k=1$ και $\beta. k=3$
4. Εφαρμόστε τον αλγόριθμο IB2 στα αρχικά δεδομένα εκπαίδευσης, επιλέγοντας αρχικά το σημείο A για μεταφορά στο Condensing Set (CS) και εξετάζοντας τα υπόλοιπα σημεία με αλφαβητική σειρά. Δώστε το τελικό CS (condensing set)
5. Σε ποια κλάση θα κατηγοριοποιηθεί το Q αν χρησιμοποιήσουμε το CS που παρήγαγε ο IB2 και ορίσουμε $\alpha. k=1$ και $\beta. k=3$
6. Εφαρμόστε τον αλγόριθμο IB2 στα δεδομένα εκπαίδευσης του ES που προέκυψε από το ερώτημα 2, επιλέγοντας αρχικά το σημείο A για μεταφορά στο CS και εξετάζοντας τα υπόλοιπα σημεία με αλφαβητική σειρά. Δώστε το τελικό CS (condensing set)
7. Σε ποια κλάση θα κατηγοριοποιηθεί το Q αν χρησιμοποιήσουμε το CS που προέκυψε στο ερώτημα 6 και ορίσουμε $k=1$
8. Εφαρμόστε τον αλγόριθμο IB2 στα αρχικά δεδομένα εκπαίδευσης, επιλέγοντας αρχικά το σημείο A και εξετάζοντας τα υπόλοιπα σημεία με αντίστροφη αλφαβητική σειρά. Δώστε το τελικό CS (condensing set)
9. Σε ποια κλάση θα κατηγοριοποιηθεί το Q αν χρησιμοποιήσουμε το CS του ερωτήματος 8 και ορίσουμε $\alpha. k=1$ και $\beta. K=3$
10. Εφαρμόστε τον αλγόριθμο CNN στα αρχικά δεδομένα εκπαίδευσης, επιλέγοντας για μεταφορά στο CS το σημείο A και εξετάζοντας τα υπόλοιπα σημεία με αλφαβητική σειρά. Δώστε το τελικό CS (condensing set)
11. Εφαρμόστε τον αλγόριθμο CNN στα δεδομένα εκπαίδευσης του ES που προέκυψε από το ερώτημα 2, επιλέγοντας για μεταφορά στο CS το σημείο A και εξετάζοντας τα υπόλοιπα σημεία με αλφαβητική σειρά. Δώστε το τελικό CS (condensing set)

ΛΥΣΕΙΣ

1. α. “μαύρος κύκλος” β. “λευκός κύκλος”.

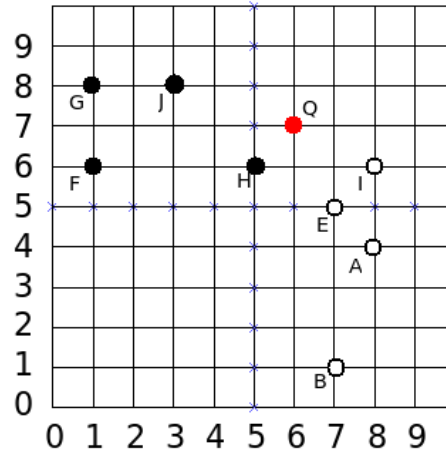
2.

εξετάζω σημείο

ES

A	A, B, C, D, E, F, G, H, I, J, K
B	A, B, C, D, E, F, G, H, I, J, K
C φεύγει	A, B, D, E, F, G, H, I, J, K
D φεύγει	A, B, E, F, G, H, I, J, K
E	A, B, E, F, G, H, I, J, K
F	A, B, E, F, G, H, I, J, K
G	A, B, E, F, G, H, I, J, K
H	A, B, E, F, G, H, I, J, K
I	A, B, E, F, G, H, I, J, K
J	A, B, E, F, G, H, I, J, K
K φεύγει	A, B, E, F, G, H, I, J

ES (ENN K=3)



3. α. “μαύρος κύκλος” β. “λευκός κύκλος”.

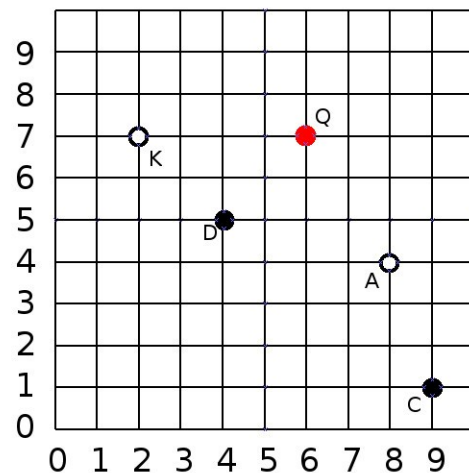
4.

εξετάζω σημείο

CS

A	A
B	A
C	A, C
D	A, C, D
E	A, C, D
F	A, C, D
G	A, C, D
H	A, C, D
I	A, C, D
J	A, C, D
K	A, C, D, K

CS (IB2)



5. α. “μαύρος κύκλος” β. “λευκός κύκλος”.

6.

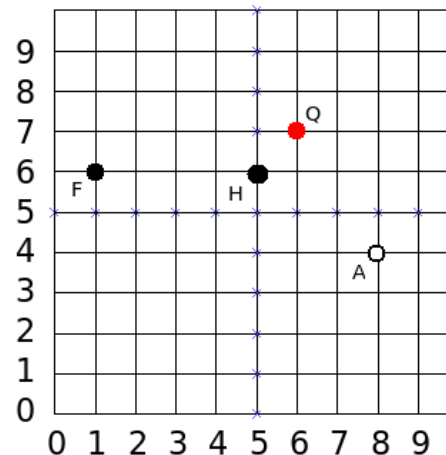
εξετάζω σημείο

CS

A	A
B	A
E	A
F	A, F
G	A, F
H	A, F, H
I	A, F, H
J	A, F, H

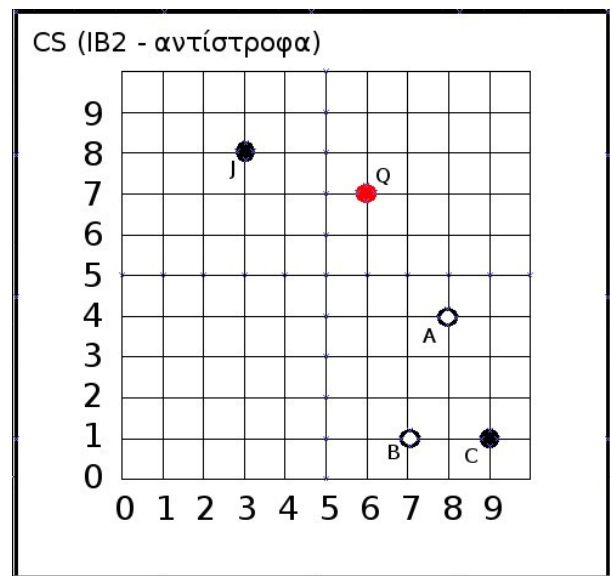
7. “μαύρος κύκλος”

ES (ENN K=3 - IB2)



8.

εξετάζω σημείο	CS
K	A
J	A
I	A, J
H	A, J
G	A, J
F	A, J
E	A, J
D	A, J
C	A, J, C
B	A, J, C, B



9. α. “μαύρος κύκλος” β. “λευκός κύκλος”.

10.

PASS 1	
TS	CS
A	A
B	A
G	A, C
D	A, C, D
E	A, C, D
F	A, C, D
G	A, C, D
H	A, C, D
I	A, C, D
J	A, C, D
K	A, C, D, K
PASS 2	
B	A, C, D, K, B
E	A, C, D, K, B
F	A, C, D, K, B, F
G	A, C, D, K, B, F, G
H	A, C, D, K, B, F, G
I	A, C, D, K, B, F, G
J	A, C, D, K, B, F, G, J
PASS 3	

Καμία μετακίνηση. Ο CNN τερματίζει

Εξαιτίας του θορύβου στα δεδομένα, πολλά στιγμιότυπα μετακινούνται λανθασμένα στο CS αφού ο CNN τα θεωρεί border items. Για παράδειγμα, αν δεν υπήρχε το στιγμιότυπο K στα δεδομένα, τα στιγμιότυπα G, F, J δεν θα αποθηκεύονταν στο CS. Για αυτό τον λόγο, οι αλγόριθμοι condensing πρέπει να εκτελούνται σε “καθαρά” (χωρίς θόρυβο) δεδομένα. Δοκιμάστε να εκτελέσετε τον CNN στα δεδομένα του ES. Θα παρατηρήσετε ότι οι μετακινήσεις είναι λιγότερες. Άρα, οδηγούμαστε σε υψηλότερα reduction rates.

11.

PASS 1

TS

CS

A

A

B

A

E

A

F

A, F

G

A, F

H

A,F,H

I

A,F,H

G

A,F,H

PASS 2

Καμία μετακίνηση. Ο CNN τερματίζει