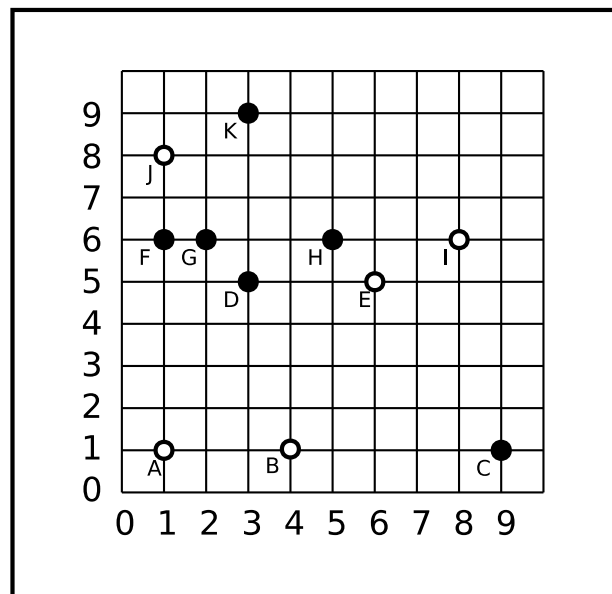


## Εργασία 2 – Data Reduction, Clustering (με το χέρι)

(Α) Στο σχήμα που σας δίνεται, φαίνονται στο Ευκλείδειο επίπεδο 11 σημεία που ανήκουν στις κλάσεις “μαύρος κύκλος” και “λευκός κύκλος” και αποτελούν το σύνολο εκπαίδευσης (TS = training set) για τον kNN. Για ευκολία έχουμε δώσει ονόματα στα σημεία (γράμματα Α ως Κ).

**Μετρική απόστασης είναι η απόσταση Manhattan (σε περίπτωση ισοπαλίας εγγύτερος είναι ο αλφαβητικά μικρότερος γείτονας).** Για ευκολία τα 11 σημεία έχουν ακέραιες συντεταγμένες και στην πράξη δεν απαιτείται να υπολογίσετε τις αποστάσεις (ο εγγύτερος γείτονας φαίνεται με το “μάτι”).



Επίσης σας δίνεται ο ψευδοκώδικας των αλγορίθμων ENN-rule και IB2.

---

### Algorithm ENN-rule

---

**Input:**  $TS, k$

**Output:**  $ES$

```
1:  $ES \leftarrow TS$ 
2: for each  $x \in TS$  do
3:    $NNs \leftarrow$  find the  $k$  nearest to  $x$  neighbors in  $TS - \{x\}$ 
4:    $majorClass \leftarrow$  find the most common class of  $NNs$ 
5:   if  $x_{class} \neq majorClass$  then
6:      $ES \leftarrow ES - \{x\}$ 
7:   end if
8: end for
9: return  $ES$ 
```

---

### Algorithm IB2

---

**Input:**  $TS$  **Output:**  $CS$

```
1:  $CS \leftarrow \emptyset$ 
2: pick an item of  $TS$  and move it to  $CS$ 
3: for each  $x \in TS$  do
4:    $NN \leftarrow$  Nearest Neighbour of  $x$  in  $CS$ 
5:   if  $NN_{class} \neq x_{class}$  then
6:      $CS \leftarrow CS \cup x$ 
7:   end if
8:    $TS \leftarrow TS - x$ 
9: end for
10: return  $CS$ 
```

---

- Εφαρμόστε τον αλγόριθμο editing ENN-rule με  $k=3$  στο TS και δώστε το ES (edited set) που θα προκύψει.
- Εφαρμόστε τον αλγόριθμο data reduction IB2 στο TS επιλέγοντας στο βήμα 2 το σημείο K και εξετάζοντας τα υπόλοιπα σημεία με αντίστροφη αλφαβητική σειρά. Δώστε το τελικό CS (condensing set).
- Προβλέψτε την κλάση του query point με συντεταγμένες (6,7) εφαρμόζοντας τον kNN με  $k=3$  διαδοχικά στα TS, ES και CS.

(B) Το παρακάτω σχήμα παρουσιάζει 10 σημεία με ακέραιες συντεταγμένες στον μονοδιάστατο χώρο.

a		b		c		d		e		f		g		h		i		j		
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Στα παρακάτω όποτε υπάρχει περίπτωση ισοβαθμίας να επιλέγετε πάντα την επιλογή στα αριστερά.

1. Εφαρμόστε Ιεραρχική συσταδοποίηση με μέτρο ομοιότητας συστάδων τη μέθοδο MIN distance (simple linkage). Δώστε τα διαδοχικά βήματα του αλγορίθμου:

ΒΗΜΑ 1: a(0) b(4) c(5) d(8) e(12) f(14) g(16) h(17) i(18) j(20)  
 ΒΗΜΑ 2: a(0) bc(4,5) d(8) e(12) f(14) g(16) h(17) i(18) j(20)  
 ΒΗΜΑ 3: a(0) bc(4,5) d(8) e(12) f(14) gh(16,17) i(18) j(20)  
 ΒΗΜΑ 4: ...

2. Εφαρμόστε Ιεραρχική συσταδοποίηση με μέτρο ομοιότητας συστάδων τη μέθοδο MAX distance (Complete linkage). Δώστε τα διαδοχικά βήματα του αλγορίθμου.

3. Εφαρμόστε τον k-means με k=3 και αρχικά centroids τα σημεία 15, 17, 19. Δώστε τα διαδοχικά βήματα του αλγορίθμου (C<sub>i</sub> είναι οι συστάδες):

ΒΗΜΑ 1: C1Centroid=15 C2Centroid=17 C3Centroid=19  
 ανάθεση σημείων: C1(abcdefg) C2(hi) C3(j)  
 ΒΗΜΑ 2: C1Centroid=8,43 C2Centroid=17,5 C3Centroid=20  
 ανάθεση σημείων: C1(abcde) C2(fghi) C3(j)  
 ΒΗΜΑ 3: ...

4. Εφαρμόστε τον DBSCAN με eps=3 και minpts=3.