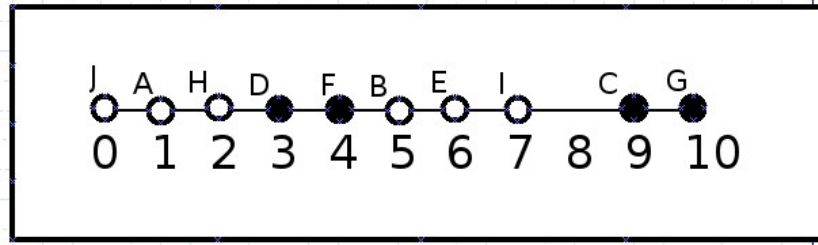


Μείωση του όγκου των δεδομένων: Σύνοψη δεδομένων (data abstraction)

Σας δίνεται το παρακάτω σύνολο δεδομένων στο μονοδιάστατο χώρο:



Επίσης σας δίνονται οι αλγόριθμοι: CJA και RSP3 σε ψευδογλώσσα:

Algorithm 6 CJA

Input: TS, n

Output: CS

```
1:  $S \leftarrow \emptyset$ 
2:  $\text{add}(S, TS)$ 
3: for  $i = 2$  to  $n$  do
4:    $C \leftarrow$  select the non-homogeneous subset  $\in S$  with the largest diameter
5:   if  $C == \emptyset$  {All subsets are homogeneous} then
6:      $C \leftarrow$  select the homogeneous subset  $\in S$  with the largest diameter
7:   end if
8:    $(S_x, S_y) \leftarrow$  divide  $C$  into two subsets
9:    $\text{add}(S, S_x)$ 
10:   $\text{add}(S, S_y)$ 
11:   $\text{remove}(S, C)$ 
12: end for
13:  $CS \leftarrow \emptyset$ 
14: for each subset  $T \in S$  do
15:    $r \leftarrow$  compute the mean item by averaging the items in  $T$ 
16:    $r.\text{label} \leftarrow$  find the most common class label in  $T$ 
17:    $CS \leftarrow CS \cup \{r\}$ 
18: end for
19: return  $CS$ 
```

Algorithm 7 RSP3

Input: TS

Output: CS

```
1:  $S \leftarrow \emptyset$ 
2:  $\text{add}(S, TS)$ 
3:  $CS \leftarrow \emptyset$ 
4: repeat
5:    $C \leftarrow$  select the subset  $\in S$  with the highest splitting criterion value
6:   if  $C$  is homogeneous then
7:      $r \leftarrow$  calculate the mean item by averaging the items in  $C$ 
8:      $r.\text{label} \leftarrow$  class of items in  $C$ 
9:      $CS \leftarrow CS \cup \{r\}$ 
10:  else
11:     $(D_1, D_2) \leftarrow$  divide  $C$  into two subsets
12:     $\text{add}(S, D_1)$ 
13:     $\text{add}(S, D_2)$ 
14:     $\text{remove}(S, C)$ 
15:  end if
16: until  $\text{IsEmpty}(S)$ 
17: return  $CS$ 
```

E

1. Να παρουσιάσετε το condensing set (CS) μετά την εκτέλεση του CJA στα δεδομένα του παραπάνω σχήματος αν ορίσουμε $n = 4$ (n : πλήθος των groups)
2. Να παρουσιάσετε το condensing set (CS) μετά την εκτέλεση του RSP1 στα δεδομένα του παραπάνω σχήματος αν ορίσουμε $n = 4$ (n : πλήθος των groups)
3. Να παρουσιάσετε το condensing set (CS) μετά την εκτέλεση του RSP1 στα δεδομένα του παραπάνω σχήματος αν ορίσουμε $n = 3$ (n : πλήθος των groups)
4. Να παρουσιάσετε το condensing set (CS) μετά την εκτέλεση του RSP3 στα δεδομένα του παραπάνω σχήματος

Λύσεις

1.

GROUP1: JAHDFB

GROUP2: EICG

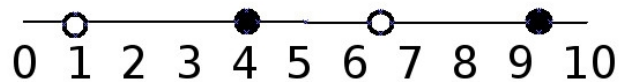
GROUP3: JAH

GROUP4: DFB

GROUP5: EI

GROUP6: CG

CJA N=4



2.

GROUP1: JAHDFB

GROUP2: EICG

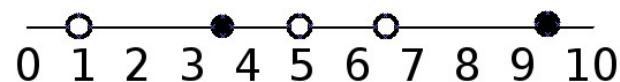
GROUP3: JAH

GROUP4: DFB

GROUP5: EI

GROUP6: CG

RSP1 N=4



3.

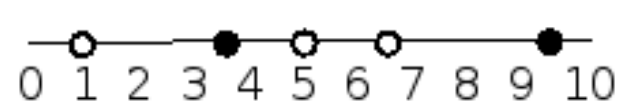
GROUP1: JAHDFB

GROUP2: EICG

GROUP3: JAH

GROUP4: DFB

RSP1 N=3



4.

GROUP1: JAHDFB

GROUP2: EICG

GROUP3: JAH

GROUP4: DFB

GROUP5: EI

GROUP6: CG

GROUP7: DF

GROUP8: B

RSP3

