

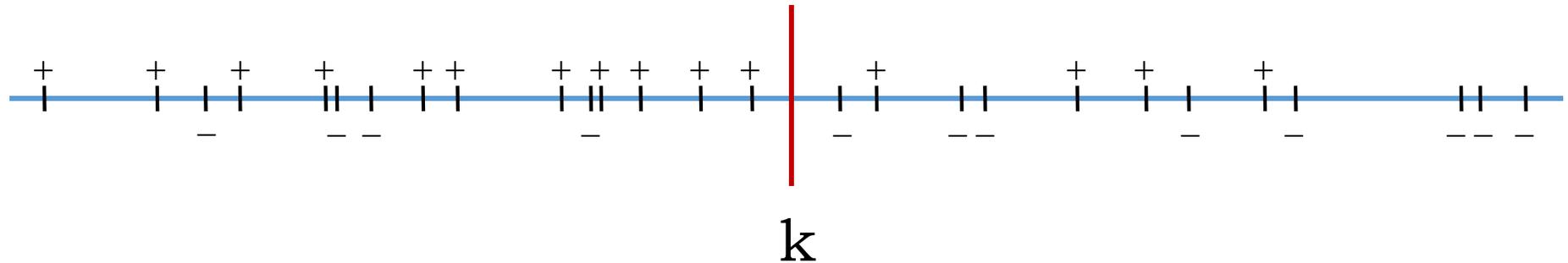


POLITÉCNICA

# Proyecto 7



- Separación de dos conjuntos de puntos unidimensionales:
  - Determinación de los valores de corte de atributos numéricos para la construcción de árboles de decisión binarios.



- Problema:

- Determinar el mejor valor de corte posible del atributo  $k$ .



POLITÉCNICA

# Proyecto 7



## □ Criterio 1:

- Dado el valor de  $k$ , llamamos

$ip$  = número de casos + a la izquierda de  $k$

$in$  = número de casos - a la izquierda de  $k$

$dp$  = número de casos + a la derecha de  $k$

$dn$  = número de casos - a la derecha de  $k$

$\epsilon_1 = in + dp$  (error para la clasificación + -)

$\epsilon_2 = ip + dn$  (error para la clasificación - +)

$\epsilon = \min(\epsilon_1, \epsilon_2)$

$ip = 11$

$in = 4$

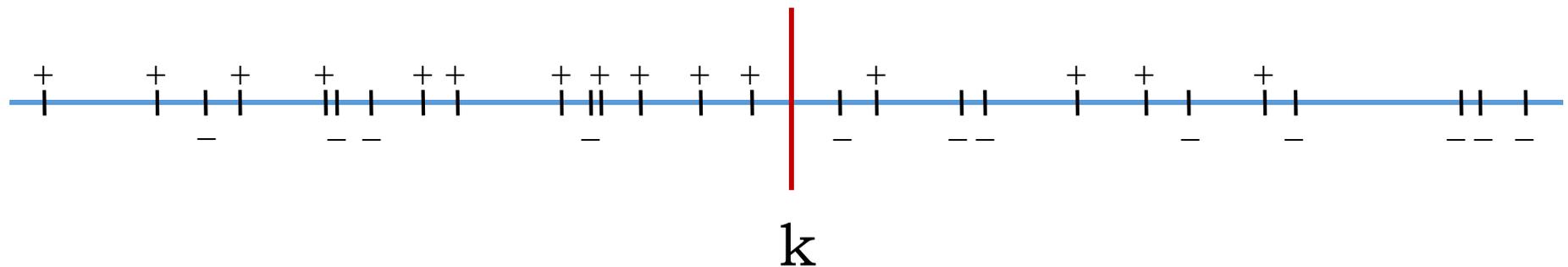
$dp = 4$

$dn = 8$

$\epsilon_1 = 8$

$\epsilon_2 = 19$

$\epsilon = 8$



## □ Objetivo-criterio 1:

- Minimizar  $\epsilon$



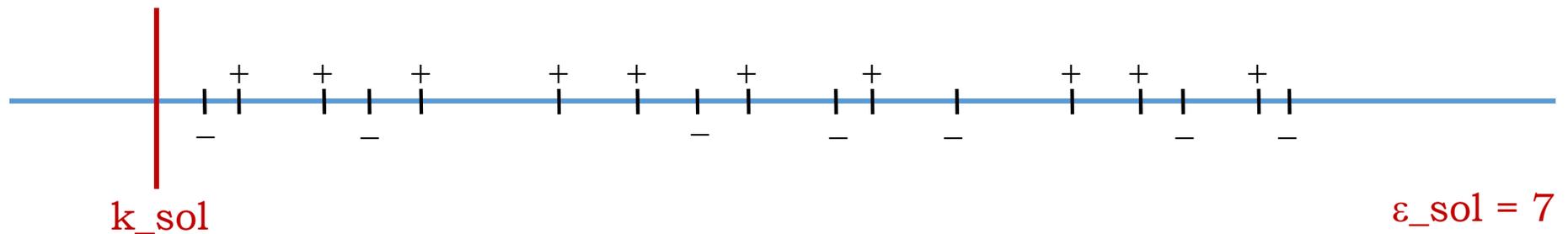
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



ip = 0  
in = 0  
dp = 10  
dn = 7  
 $\epsilon_1 = 10$   
 $\epsilon_2 = 7$   
 $\epsilon = 7$



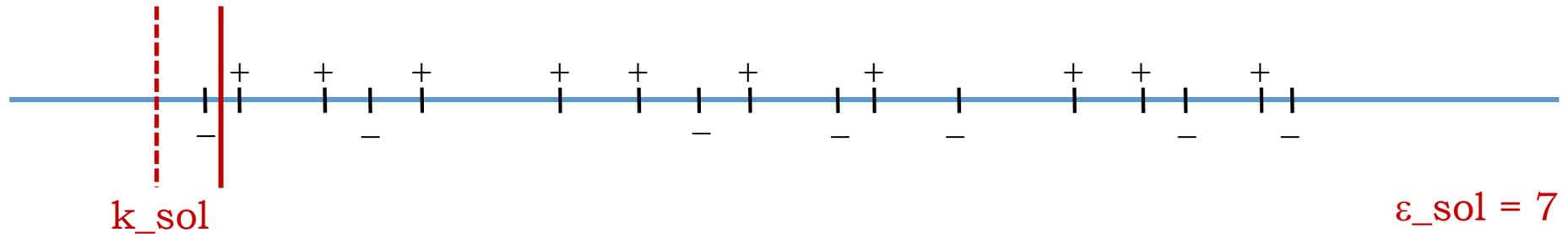
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



ip = 0 0  
in = 0 1  
dp = 10 10  
dn = 7 6  
 $\varepsilon_1$  = 10 11  
 $\varepsilon_2$  = 7 6  
 $\varepsilon$  = 7 6



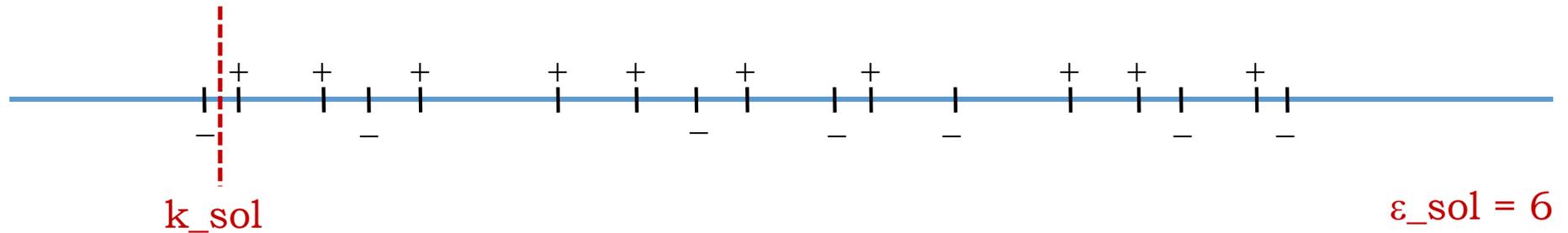
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



```
ip = 0 0
in = 0 1
dp = 10 10
dn = 7 6
ε1 = 10 11
ε2 = 7 6
ε = 7 6
```



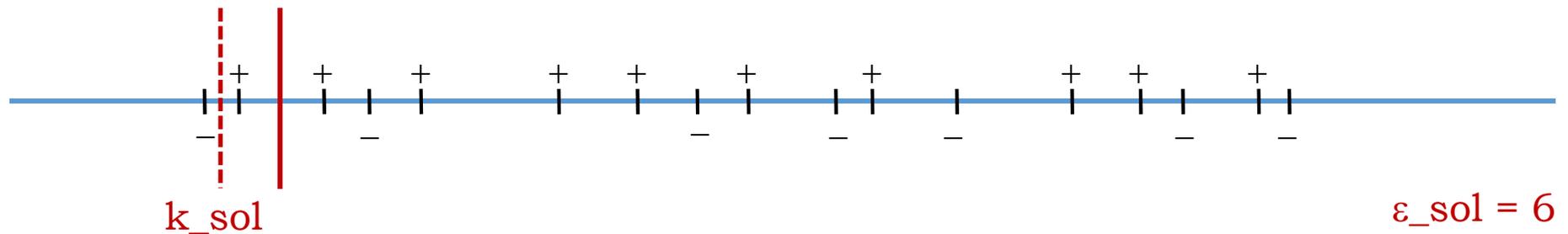
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



```
ip = 0 0 1
in = 0 1 1
dp = 10 10 9
dn = 7 6 6
ε1 = 10 11 10
ε2 = 7 6 7
ε = 7 6 7
```



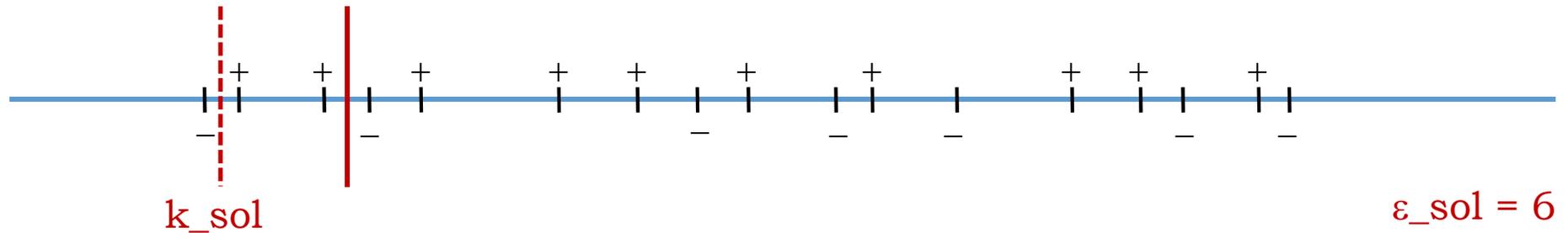
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



```

ip = 0 0 1 2
in = 0 1 1 1
dp = 10 10 9 8
dn = 7 6 6 6
ε1 = 10 11 10 9
ε2 = 7 6 7 8
ε = 7 6 7 8

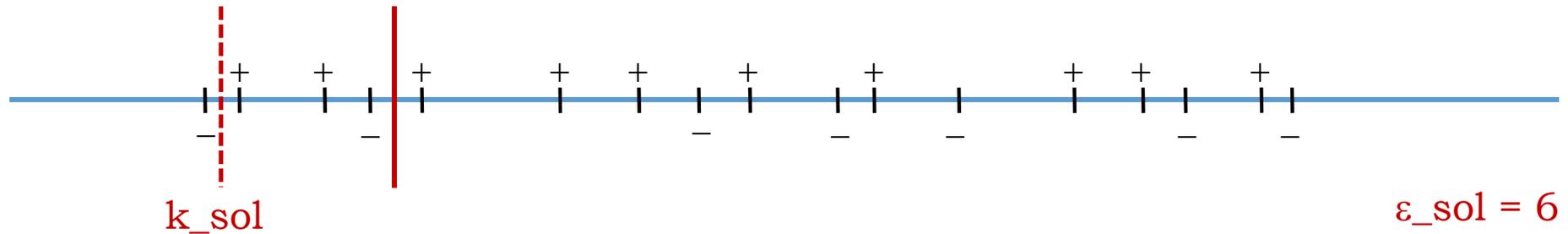
```



# Proyecto 7

## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



ip	=	0	0	1	2	2
in	=	0	1	1	1	2
dp	=	10	10	9	8	8
dn	=	7	6	6	6	5
$\epsilon_1$	=	10	11	10	9	10
$\epsilon_2$	=	7	6	7	8	7
$\epsilon$	=	7	6	7	8	7



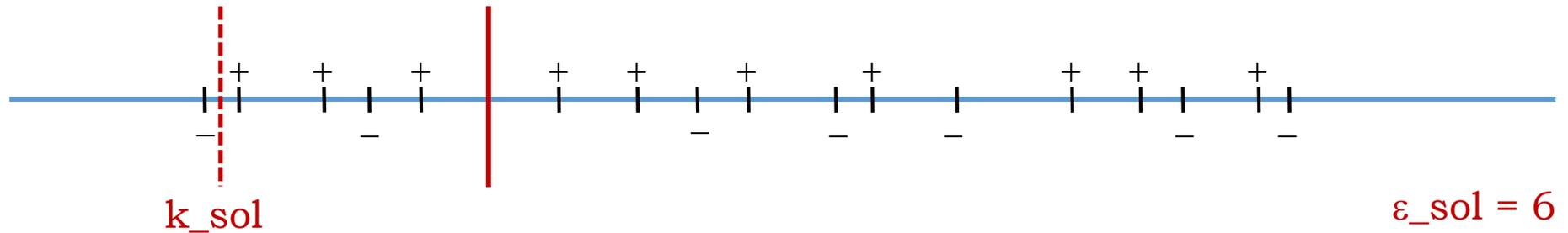
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



ip	=	0	0	1	2	2	3
in	=	0	1	1	1	2	2
dp	=	10	10	9	8	8	7
dn	=	7	6	6	6	5	5
$\epsilon_1$	=	10	11	10	9	10	9
$\epsilon_2$	=	7	6	7	8	7	8
$\epsilon$	=	7	6	7	8	7	8



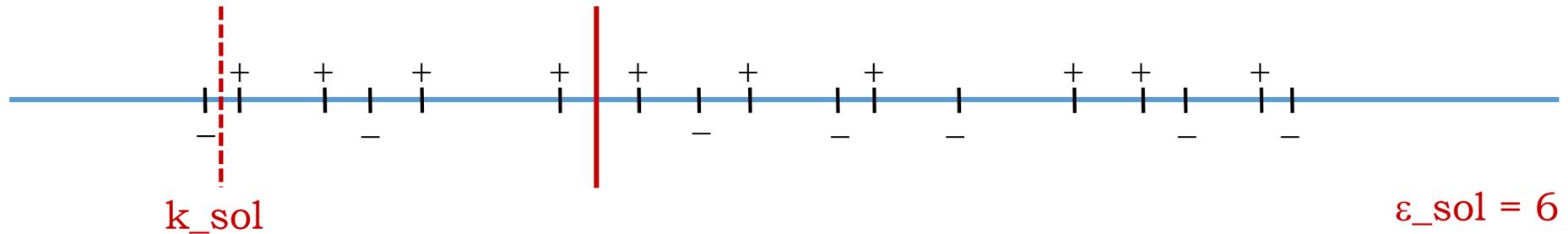
POLITÉCNICA

# Proyecto 7



## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



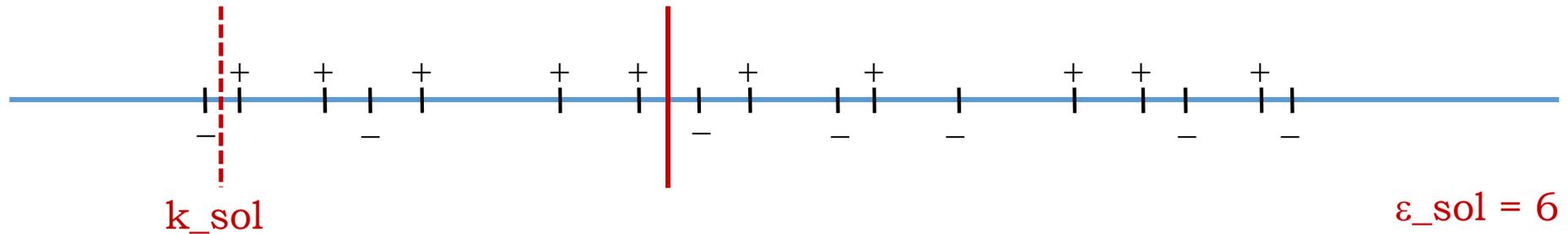
ip	=	0	0	1	2	2	3	4
in	=	0	1	1	1	2	2	2
dp	=	10	10	9	8	8	7	6
dn	=	7	6	6	6	5	5	5
$\epsilon_1$	=	10	11	10	9	10	9	8
$\epsilon_2$	=	7	6	7	8	7	8	9
$\epsilon$	=	7	6	7	8	7	8	8



# Proyecto 7

## □ Ejemplo algoritmo-criterio 1:

- Damos valores a  $k$  desde  $\min-1$  hasta  $\max+1$ , pasando por todos los puntos medios de cada par de valores consecutivos del atributo



ip	=	0	0	1	2	2	3	4	5	
in	=	0	1	1	1	2	2	2	2	
dp	=	10	10	9	8	8	7	6	5	
dn	=	7	6	6	6	5	5	5	5	.....
$\epsilon_1$	=	10	11	10	9	10	9	8	7	
$\epsilon_2$	=	7	6	7	8	7	8	9	10	
$\epsilon$	=	7	6	7	8	7	8	8	7	



POLITÉCNICA

# Proyecto 7



## □ Algoritmo-criterio 1:

- Ordenamos los valores del atributo en orden creciente:  $A[0] \rightarrow A[n-1]$
- Inicializamos  $k = A[0] - 1$  y  $k\_sol = k$
- Calculamos  $ip$ ,  $in$ ,  $dp$ ,  $dn$ ,  $\varepsilon_1$ ,  $\varepsilon_2$  y  $\varepsilon$  e inicializamos  $\varepsilon\_sol = \varepsilon$
- Para  $j$  desde 0 hasta  $n-2$ :
  - Asignamos  $k = (A[j] + A[j+1])/2$
  - Si  $AO[j] = +$ :  $ip += 1$ ,  $dp -= 1$ ,  $\varepsilon_1 -= 1$ ,  $\varepsilon_2 += 1$
  - En otro caso:  $in += 1$ ,  $dn -= 1$ ,  $\varepsilon_1 += 1$ ,  $\varepsilon_2 -= 1$
  - Recalculamos  $\varepsilon$
  - Si  $\varepsilon < \varepsilon\_sol$ :  $\varepsilon\_sol = \varepsilon$ ,  $k\_sol = k$
- Asignamos  $k = A[n-1] + 1$
- Si  $AO[n-1] = +$ :  $ip += 1$ ,  $dp -= 1$ ,  $\varepsilon_1 -= 1$ ,  $\varepsilon_2 += 1$
- En otro caso:  $in += 1$ ,  $dn -= 1$ ,  $\varepsilon_1 += 1$ ,  $\varepsilon_2 -= 1$
- Recalculamos  $\varepsilon$
- Si  $\varepsilon < \varepsilon\_sol$ :  $\varepsilon\_sol = \varepsilon$ ,  $k\_sol = k$
- Devolver( $k\_sol$ ,  $\varepsilon\_sol$ )



POLITÉCNICA

# Proyecto 7



□ Input: archivo con el siguiente formato

m <número de atributos (de tipo entero salvo el último que es el objetivo)>  
n <número de ejemplos>  
v <valor del primer atributo del primer ejemplo>  
...  
v <valor del último atributo del primer ejemplo>  
...  
v <valor del primer atributo del último ejemplo>  
...  
v <valor del último atributo del último ejemplo>



POLITÉCNICA

# Proyecto 7



□ Output: archivo con el siguiente formato

k <valor de corte del primer atributo>

...

k <valor de corte de penúltimo atributo>