Proyecto 8

Clustering unidimensional:

- Dado un conjunto de n puntos en orden creciente, $x_1, x_2, \ldots x_n$, y un entero $k \geq 1$, asignar cada uno de los puntos x_i a un subconjunto S_j, $1 \leq j \leq k$, de forma que se minimice

$$D = \sum_{1\leq j\leq k} \left(\sum_{x_i \in S_j} (x_i - c_j)^2 \right) \qquad c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i, \quad 1 \leq j \leq k$$

Propiedad básica:

- Si ordenamos en orden creciente los puntos medios de los subconjuntos S_j, $c_1, c_2, \ldots c_k$, entonces se verifica

$$\max(S_j) \leq \min(S_{j+1}) \qquad 1 \leq j < k$$

Proyecto 8

- Algoritmo (programación dinámica):
  - Calcular la solución óptima D[i,m] para todo $1 \leq i \leq n$ y $1 \leq m \leq k$

$$D[i,m] = \min_{m \leq j \leq i} \{ D[j-1,m-1] + d(x_j,\ldots,x_i) \}$$

$d(x_j,\ldots x_i)$ es la suma de distancias al cuadrado desde $x_j,\ldots,x_i$ a su media

Nota: $D[i,m] = 0 \quad \text{si} \quad i \leq m$

$$D[i,1] = \sum_{1 \leq j \leq i} (x_j - c)^2 \quad c = \frac{1}{i} \sum_{1 \leq j \leq i} x_j, \quad i \geq 2$$

Proyecto 8

❑ Cálculo de los subconjuntos S_j:

- A lo largo del algoritmo calcular la matriz B[i,m] que contiene el índice j de elemento más pequeño x_j en el subconjunto m (S_m)

$$B[i,m] = \underset{m \le j \le i}{\mathrm{indmin}} \{ D[j-1,m-1] + d(x_j,\ldots,x_i) \}$$

❑ Salida del algoritmo:

- Valor mínimo de la función a optimizar D[n,k]
- Subconjuntos solución S_j $1 \le j \le k$:

$$S_k = \{ x_{j1}, \ldots, x_n \} \quad \text{con} \quad j1=B[n,k]$$

$$S_{k-1} = \{ x_{j2}, \ldots, x_{j1-1} \} \quad \text{con} \quad j2=B[j1-1,k-1]$$

$$S_{k-2} = \{ x_{j3}, \ldots, x_{j2-1} \} \quad \text{con} \quad j3=B[j2-1,k-2]$$

....

Proyecto 8

- **Input:** archivo con el siguiente formato

k	<número de subconjuntos>
n	<número de puntos>
x	<primer punto>
…	
x	<último punot>

- **Output:** archivo con el siguiente formato

D	<valor óptimo de la función objetivo>
j	<índice del primer punto del primer subconjunto>
…	
j	<índice del primer punto del último subconjunto>