



POLITÉCNICA

Proyecto 9



□ Bottom-up clustering:

Dado un conjunto S de n puntos en \mathbb{R}^d , x_1, x_2, \dots, x_n , construir un árbol binario con raíz T tal que:

- Los vértices de T son clusters de S .
- Las hojas de T son $\{x_1\}, \{x_2\}, \dots, \{x_n\}$.
- El cluster de un vértice es la unión de los clusters de los vértices hijos
- La raíz de T es $\{x_1, x_2, \dots, x_n\}$.

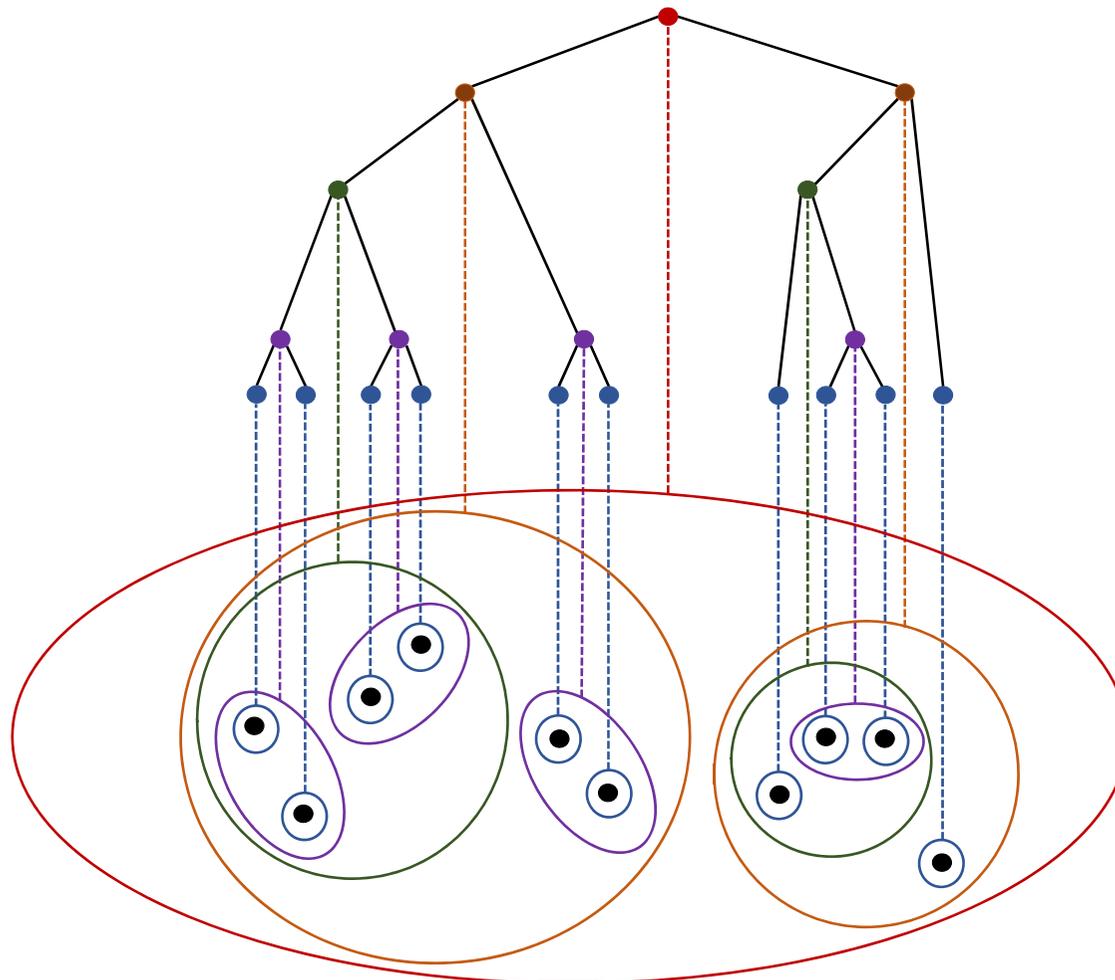


POLITÉCNICA

Proyecto 9



□ Ejemplo:





POLITÉCNICA

Proyecto 9



□ Reglas para la unión de clusters:

- Inicialmente los clusters activos son $\{x_1\}, \{x_2\}, \dots, \{x_n\}$.
- Cuando unimos dos clusters activos ($C=A \cup B$), estos dejan de serlo y pasa a ser activo el cluster resultante de la unión.
- Se eligen para la unión la pareja de clusters activos que minimiza la siguiente función:

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b) \quad \text{donde} \quad d(a, b) = \sqrt{\sum_i (a_i - b_i)^2}$$



POLITÉCNICA

Proyecto 9



□ Input: archivo con el siguiente formato

n <número de puntos>
d <dimensión del espacio \mathbb{R}^d >
 x_1 <primera coordenada del primer punto>
...
 x_d <última coordenada del primer punto>
...
 x_1 <primera coordenada del último punto>
...
 x_d <última coordenada del último punto>



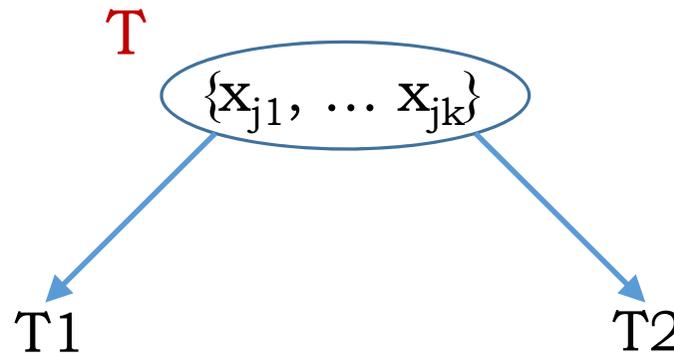
POLITÉCNICA

Proyecto 9



□ Output: cadena representando el árbol resultante

Definición recursiva:



$Salida(T) = [[j_1, \dots, j_k], Salida(T1), Salida(T2)]$