

Sesión 3.1: Variantes de k-d Tree

CS3102 EDA

Índice

1. Median-split k-d tree
2. Adaptive k-d tree
3. Fair-split k-d tree



1. Median-split k-d Tree

Median-split *k-d Tree*

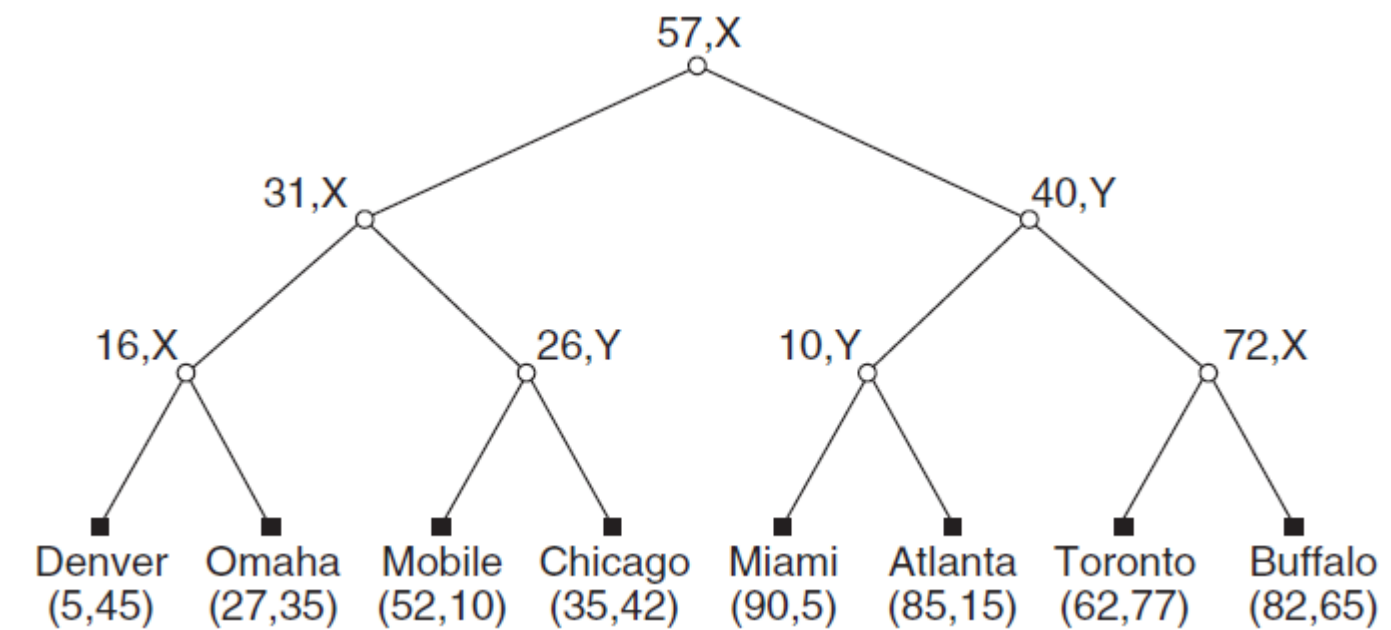
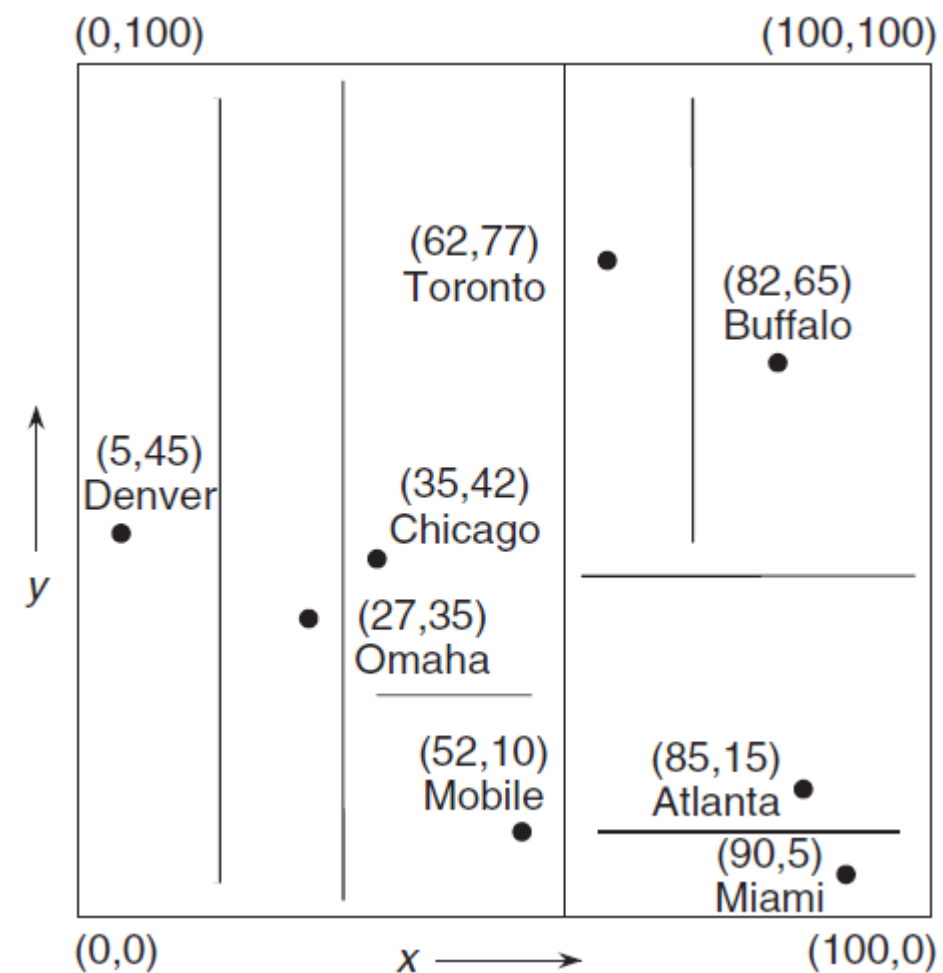
¿Si conocen los puntos de antemano, como insertarían los puntos?

Median-split *k-d Tree*

- Se **construye** dividiendo recursivamente los datos a lo largo de **la dimensión con el mayor rango** (la diferencia entre los valores máximo y mínimo).
- El **split** se realiza la **mediana** de dimensión seleccionada, garantizando que aproximadamente la mitad de los puntos se encuentren a cada lado de la división.
- Este método garantiza una estructura de árbol equilibrada.
- Puede no ser muy eficaz si la distribución de los datos está muy sesgada.

2. Adaptive k-d Tree

Adaptive k -d Tree



Adaptive k -d Tree

- Se **construye** dividiendo recursivamente los datos a lo largo de la dimensión con la **varianza más alta**.
- El **split** se elige de forma que minimice **la suma de las varianzas** de las dos subregiones resultantes.
- Este método adapta la estructura del árbol a la distribución de los datos

Adaptive *k-d Tree*

Rendimiento en
consultas

Suele funcionar mejor con datos muy sesgados o agrupados, ya que se adapta a la distribución de los datos.

Complejidad

Requiere una mayor carga computacional, ya que implica el cálculo de la varianza de cada dimensión y la determinación de la división óptima.

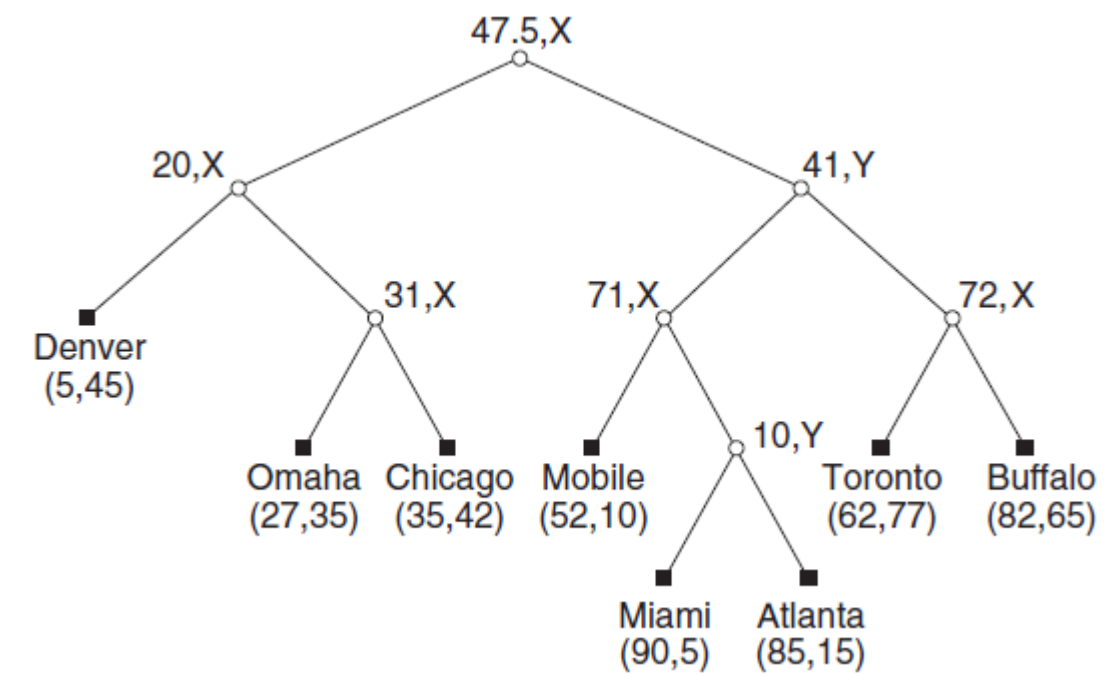
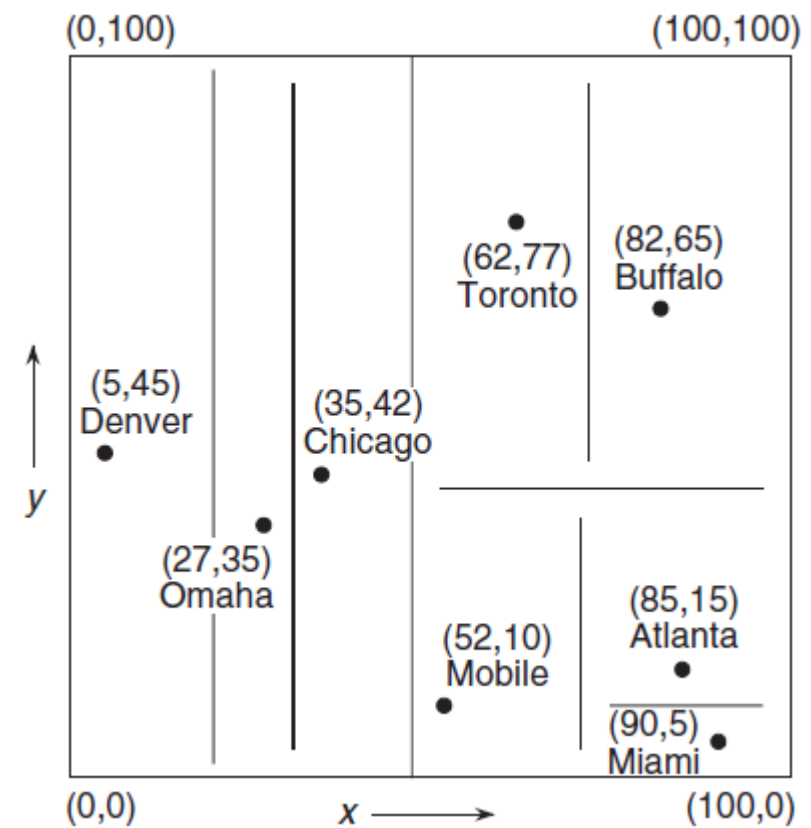
Median-split *k-d Tree*

Garantiza una estructura de árbol equilibrada, lo que asegura un rendimiento predecible en el peor de los casos. Sin embargo, puede no ser tan eficaz como el Adaptive k-d Tree cuando se manejan datos muy sesgados o agrupados.

Su construcción es más sencilla y rápida, ya que sólo requiere encontrar la mediana a lo largo de la dimensión con el mayor rango.

3. **Fair-split** k-d Tree

Fair-split k -d Tree



Fair-split *k-d Tree*

Elegir la dimensión del split

La mayor varianza o el mayor rango.

Elegir la posición del split

Se elije un punto que divida los datos en dos subconjuntos con una **cardinalidad** aproximadamente igual, al tiempo que se intenta minimizar la **suma de las varianzas**.

$$\mathcal{L} = |A - B| + \sigma^2(A) + \sigma^2(B)$$

Objetivo: $\min \mathcal{L}$



INGENIERIA
MECATRÓNICA

BIÓINGENIERÍA

INGENIERIA
CIENCIA DE
LA COMPUTACIÓN

INGENIERIA
AMBIENTAL

INGENIERIA
ENERGÍA

INGENIERIA
INDUSTRIAL

INGENIERIA
ELECTRÓNICA



UTEC

UNIVERSIDAD DE INGENIERÍA
Y TECNOLOGÍA

