Unsupervised Binning

Data Discretisation

Binning

- Unsupervised binning methods transforms numerical variables into categorical counterparts but do not use the target (class) information.
- Types of Binning
  - Equal Width
  - Equal Frequency

Equal Width Binning

- *The algorithm divides the data into k intervals of equal size. The width of the intervals is :*
- *w = (max-min)/k*
- *min+w, min+2w, ... , min+(k-1)w*

Binning By Equal-Width

w = (max-min)/k

min+w, min+2w, … , min+(k-1)w

Binning By Equal-Width

5,10,11,13,15,35,50,55,72,92,204,215,

- width=max−min
- number of bins= 3
- W= (215−5)/3= 70.

The equal width partition

- *min+w*
- 70+5=75 (from 5 to 75)=Bin 1
  5,10,11,13,15,35,50,55,72
- *min+2w= 5 + 140 = 145*
- 70+75=145 (from 75 to 145)=Bin 2:92

- 70+145=215 (from 145 to 215)=Bin 3:204,215

w = (max-min)/k

min+w, min+2w, ... , min+(k-1)w

Equal Frequency Binning

- The algorithm divides the data into k groups which each group contains approximately same number of values.
- For the both methods, the best way of determining k is by looking at the histogram and try different intervals or groups.

5,10,11,13,15,35,50,55,72,92,204,215, 300

13

Bin 1: 5,10,11 [-, 11)

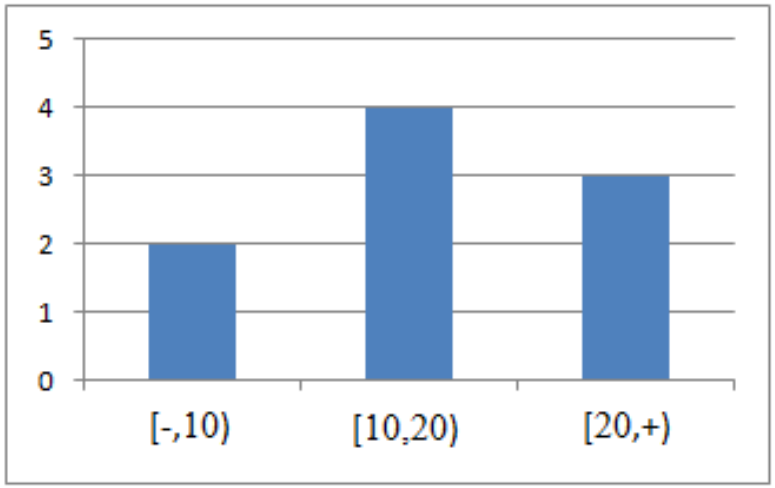
Bin 2: 13,15,35 [11, 35)

Bin 3: 50, 55, 72 [35, 72)

Bin 4: 92,204,215,300 [72, +)

- **Data :** 0, 4, 12, 16, 16, 18, 24, 26, 28
- **Equal width**
  - Bin 1: 0, 4 [-,10)
  - Bin 2: 12, 16, 16, 18 [10,20)
  - Bin 3: 24, 26, 28 [20,+)
- **Equal frequency**
  - Bin 1: 0, 4, 12 [-, 14)
  - Bin 2: 16, 16, 18 [14, 21)
  - Bin 3: 24, 26, 28 [21,+)

Equal width



Equal frequency