How to hande noisy Data →

① Binning :→

Q1. → Suppose a group of sales price records has been sorted as follows :- 6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232, Partition them into three bins by equal - frequency (Equi - depth) partitioning method. Perform data smoothing by bin mean.

⟹

① Sort the given data
6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232

② Partition the data into equal frequency bin of size = 4

Bin1 = 6, 9, 12, 13
Bin2 = 15, 25, 50, 70
Bin 3 = 72, 92, 204, 232

③ Calculate the arithmetic mean of each bin
Bin 1 = 10
Bin 2 = 40
Bin 3 = 150

④ Replace each value in the bin with it's respective arithmetic mean :-

Bin 1:- 10, 10, 10, 10
Bin 2 :- 40, 40, 40, 40
Bin 3:- 150, 150, 150, 150

Q2. for the given attribute AGE values:-
16, 16, 180, 4, 12, 24, 26, 28 apply following Binning technique for smoothing the noise:-

(i) Bin Medians (ii) Bin Boundaries
(iii) Bin means

⟹

- Sort the age in ascending order:-
  4, 12, 16, 16 24, 26, 28, 180
- Partition into (equal-depth) bins = (N=2)
- Bin 1:- 4, 12, 16, 16

Bin 2:- 24, 26, 28, 180

(i) Smoothing by Bin Medians:-

Replace each value by bin median
Bin 1:- 14, 14, 14, 14 (median = (12+16)/2 = 14

Bin2:- (26+28)/2 = 27
Bin2 = 27, 27, 27, 27

(ii) Smoothing by bin means:→

Replace each value of bin with it's mean value.

Bin 1:- mean = 12 (12, 12, 12, 12)

Bin2:- mean = 64.5 ⟹ (64.5, 64.5, 64.5, 64.5)

(iii) Smoothing by bin boundaries:-

In this method the min & max. values of the bean boundaries is found and each value is replaced with it's nearest value either min or max.

Bin 1:- 4, 16, 16, 16
Bin 2:- 24, 24, 24, 180

Different approches of binning:-

a) Equal_width (distance) partitioning

⇒ bin_width:⇒ (max. value − min. value)/N

b) Equal-depth (frequency) partitioning or Equal ~~high~~ height binning:⇒

⇒ The entire range is divided into 'N' intervals, each containing approximately the same no. of samples.