



Print Discussion Assignment 1: Binning

FUNDAMENTALS OF DATA SCIENCE Discussion Assignments Week 3 Discussion Assignment 1: Binning



Discussion Assignment 1: Binning

Created by Sadiya Amreen on Oct 27, 2022 5:50 PM [Subscribed](#)

Given Data: 18, 8, 22, 10, 12, 5, 4, 32, 2, 9, 16, 25, 26, 28

Step 1: Sort the Data in Ascending Order.

Data in Order: 2, 4, 5, 8, 9, 10, 12, 16, 18, 22, 25, 26, 28, 32

Step 2: Dividing the Data into Equal width bins.

$$W_{bin} = \frac{B-A}{N_{bins}}$$

W_{bin} = Width of the bins , A = highest value of the attribute = 32 , B = lowest value of the attribute = 2 , N = No. of bins = 3 (User Defined)

$$W_{bin} = (32 - 2) / 3 = 30 / 3 = 10$$

The Bins are divided into Equal widths.

$$B1 = [2, 12] = [2, 4, 5, 8, 9, 10, 12]$$

$$B2 = [12, 22] = [16, 18, 22]$$

$$B3 = [22, 32] = [25, 26, 28, 32]$$

Step 3: Data Smoothing by bin Median.

Calculating the median of each bin.

if n is even : median = $n/2$

if n is odd : median = $n/2$ and round off the value

so, $B1 = [2, 4, 5, 8, 9, 10, 12]$, $n = 7$, median = $7/2 = 3.5 = 4$ th value of the $B1 = 8$.

so, $B2 = [16, 18, 22]$, $n = 3$, median = $3/2 = 1.5 = 2$ nd value of the $B2 = 18$.

so, $B3 = [25, 26, 28, 32]$, $n = 4$, median = $4/2 = 2 = 2$ nd value of the $B2 = 26$.

Step 4: Replace bin values with Median values.

$$B1 = [8, 8, 8, 8, 8, 8, 8]$$

$$B2 = [18, 18, 18]$$

$$B3 = [27, 27, 27, 27]$$

New list of Data after Smoothing: [8, 8, 8, 8, 8, 8, 8, 18, 18, 18, 27, 27, 27, 27]

Part 2 : Normalization

Data in Order: 2, 4, 5, 8, 9, 10, 12, 16, 18, 22, 25, 26, 28, 32

Min-max Normalization : Range (10-20):

$$v' = \frac{v - \min_A}{\max_A - \min_A} \left((\text{new_max}_A) - (\text{new_min}_A) \right) + \text{new_min}_A$$

newminA = 10 , newmaxA= 20 , minA= 2 , maxA = 32

$$V' = \frac{2-2}{32-2} ((20) - (10)) + 10 = 10$$

$$V' = \frac{4-2}{32-2} ((20) - (10)) + 10 = 10.6$$

$$V' = \frac{5-2}{32-2} ((20) - (10)) + 10 = 11$$

Normalized data (Min-Max) :

10, 10.6, 11, 12, 12.3, 12.6, 13.3, 14.6, 15.3, 16.6, 17.6, 18, 18.6, 20

Value getting mapped to 0 is 2

min value after normalization = 10

max value after normalization = 20

z-score normalization/ standardization:

$$z = \frac{x - \mu}{\sigma}$$

Mean = μ = Sum of observations / total number of observations = 15.5

$$\text{Standard deviation} = \sigma = \sqrt{\frac{\sum (x - \mu)^2}{n-1}} = 9.780$$

Normalized data (Z-Score) :

(-1.38, -1.17 -1.07,-0.76, -0.66, -0.56,-0.35, 0.05, 0.25 ,0.66,0.97, 1.07,1.27,1.68)

Value getting mapped to 0 is None

min value after normalization = -1.38

max value after normalization = 1.68

Decimal scaling normalization: [-1,1]

Data in Order: 2, 4, 5, 8, 9, 10, 12, 16, 18, 22, 25, 26, 28, 32

$$v' = \frac{v}{10^j}$$

v= datapoints

j=where j is the smallest integer such that $\text{Max}(|v|) < 1$, In this case j=2

v' = normalized value

Normalized data (Decimal Scaling) :

(0.02 , 0.04, 0.05, 0.09, 0.10, 0.12, 0.16, 0.18, 0.22, 0.25, 0.26, 0.28, 0.32)

Value getting mapped to 0 is None

min value after normalization = 0.02

max value after normalization = 0.32

No items found.