* Solved Problems on Normalization Techniques. *

MU June-21 (10M).

① Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order): 13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72

i) Use min-max normalization to transform the value 45 for age onto the range [0.0, 1.0]

ii) Use Z-score normalization to transform the value 45 for age, where the standard deviation of age is 20.64 years.

Solution :-

By min-max normalization,

$$V' = \frac{V - min_A}{max_A - min_A} (new\_max_A - new\_min_A) + new\_min_A$$

where A is attribute data. Here, A is the attribute age.

$min_A = 13$ $max_A = 72$.

V' = its new value, $V = 45$

$new\_min_A = 0$ and $new\_max_A = 1.0$ ..... {Given in problem statement}

$$v' = \frac{45 - 13}{72 - 13} (1.0 - 0.0) + 0.0$$

$$v' = \frac{45-13}{72-13} (1-0) + 0$$

$$v' = \frac{32}{59} (1) = 0.5423$$

Hence the value 45 for age is transformed to 0.5423.

$$\text{Z score} = \frac{\text{Old value} - \text{Mean}}{\text{standard deviation}}$$

OR

$$v' = \frac{v - \mu_A}{\sigma_A}$$

here, Standard deviation is given in problem.

i.e $\sigma_A = 20.64$.

Now calculate the mean first. $\mu_A = 35.13$.

$$v' = \frac{45 - 35.13}{20.64}$$

$$v' = 0.478$$

Hence the value 45 for the age is transformed to 0.478 using Z-score transformation.

* Min-max normalization *

ex. Suppose the income range from $10,000 to $95000 is normalized to [0.0, 1.0]. By min-max normalization, a value of $64,300 is transformed to ?

$$v' = \frac{v - min_A}{max_A - min_A} (new\_max_A - new\_min_A) + new\_min$$

$$= \frac{64300 - 10000}{95000 - 10000} (1 - 0) + 0$$

$$= \frac{54300}{85000} (1)$$

$$= 0.6388$$

Hence the transformed value of 64,300 will be 0.6388

Z-score transformation

$$v' = \frac{v - \mu A}{\sigma A}$$

ex. if mean salary is $54000 and standard deviation is $16000. Then Z-score value of salary $73600 will be ?

Sol^n ⇒

$$v' = \frac{v - \mu A}{\sigma A} = v' = \frac{73600 - 54000}{16000}$$

$$= \frac{19600}{16000}$$

$$= 1.225$$

* Decimal scaling *.

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

where, j is the smallest integer such that $\max(|v_i'|) < 1$.

Ex. Let the input data be : -10, 201, 301, -401, 501, 601, 701.

To normalized the above data.

Step 1 :- maximum absolute value in given data (m) : 701

Step 2 :- So now put $j = 3$, means divide the given data by 1000. i.e ($j = 3$).

Hence, $\frac{-10}{1000}, \frac{201}{1000}, \frac{301}{1000}, \frac{-401}{1000}, \frac{501}{1000}, \frac{601}{1000}, \frac{701}{1000}$

Hence the normalized data is,

-0.01, 0.201, 0.301, -0.401, 0.501, 0.601, 0.7