Submission Details

Grade: **2 / 2**

Show Rubric

Practice Normalizing Data Using Python [Try-It Activity 4.1]

Roy Nunez submitted Apr 19 at 12:48am

Re-submit Assignment

Live Support

In this activity, we encountered a challenge where one feature, Col A, was significantly larger than the others, making it difficult to compare them directly. To address this, we explored three normalization techniques: Max Scaling, Min-Max Scaling, and Standardization.

The final bar graph of the standardized data, generated using **df_z_scaled.plot(kind='bar')**, differs significantly from the graphs produced by Max Scaling and Min-Max Scaling. Standardization centers each feature's values around zero and scales them based on the standard deviation, resulting in all features having a standard deviation of one. This method allows the bar graph to show variations both above and below the zero line, indicating how each value deviates from the mean.

In contrast, Max Scaling normalizes each feature to a range between 0 and 1, but does not center the data around zero, thus all features appear non-negative. Min-Max Scaling also normalizes each feature to a range between 0 and 1, effectively compressing the range but not centering the data, leading to a uniform positive range across all features.

2. Can you think of a scenario in your current workplace where one of these normalization techniques would have been helpful? If you cannot think of a current scenario, think of a project you worked on in the past where these techniques could have been applied. How will you use these techniques in the future?

I think in a past mentioned project involving customer segmentation for marketing analysis, these normalization techniques would be very useful. While I think the standardization technique was impressive in terms of the activity example I think min-max scaling could have been very useful for our use case. We could have scaled the income levels, transaction frequencies, and have the engagement scores from 0 to 1.

For example, the income levels can range from thousands to millions. Using the formula for min-max scaling

$$\text{Scaled Value} = (\text{Original Value} - \text{Min Value}) / (\text{Max Value} - \text{Min Value})$$

in a scenario where the minimum income is $20,000 and the maximum is $500,000, and income of 170,000 will be transformed to .31

$$(170{,}000 - 20{,}000) / (500{,}000 - 20{,}000) = 0.31$$

For transactions, if the minimum is 1 and the maximum is 100, a frequency of 50 would be scaled to 0.49.

Following this normalization technique of min-max scaling would have made it easier to profile cust[illegible]le scale without allowing one metric such as income, dominate and dis[illegible]analysis.In this way each attribute would have contributed equally to the

segmentation process, allowing for improved insights and more effective marketing strategies tailored to

Add a Comment:

Attach File

Save