



Dash



All



Articles



Videos



Quiz

Outlier Removal using Percentile

Understanding Percentile-Based Outlier Removal:

Percentiles are a measure that indicates a specific value below which a given percentage of observations fall. For instance, the 25th percentile (Q1) indicates the value below which 25% of the data points lie. By utilizing percentiles, we can establish thresholds that define what values are considered normal and what are potential outliers.

Python3

```
import numpy as np

# Generate example data
data = np.array([23, 25, 22, 27, 21, 24, 26, 100, 23, 28, 22, 29])

# Define percentiles
percentile_lower = 10
percentile_upper = 90

# Calculate the percentile values
lower_bound = np.percentile(data, percentile_lower)
upper_bound = np.percentile(data, percentile_upper)

# Identify outliers
outliers = [x for x in data if x < lower_bound or x > upper_bound]

print("Original Data:", data)
print("Outliers detected using Percentiles:", outliers)
```

Explanation of the Code:

- 1. We import the NumPy library, which provides functions for numerical operations in Python.
- 2. We create an example dataset called `data` containing a mixture of normal and outlier values.
- 3. We define the desired percentiles (`percentile_lower` and `percentile_upper`) to establish the boundaries for outlier detection.
- 4. We calculate the corresponding percentile values using the `percentile` function from NumPy.
- 5. We identify outliers by checking if a data point falls outside the defined percentile boundaries.
- 6. Finally, we print both the original dataset and the detected outliers.

Advantages of Percentile-Based Outlier Removal:

- 1. **Robustness:** Percentile-based outlier removal is less sensitive to extreme values, making it suitable for skewed or non-normal distributions.
- 2. **Flexibility:** By adjusting the percentile thresholds, you can control the strictness of outlier removal to match the characteristics of your data.

Mark as Read

Report An Issue

If you are facing any issue on this page. Please let us know.

<< Prev

Next >>