

What are outliers / outliers in Machine Learning

what are outliers?

outliers are those data points that are significantly different from the rest of the dataset. They are often abnormal.

when is outlier dangerous?

observations that skew the data distribution.

Effect of outliers on ML algorithms

→ linear regression,
logistic regression,
Ada boost, deep learning
weight based models

How to treat outliers?

How to detect outliers?

Techniques for outliers Detection

Trimming

if outliers are large in number then data become thin. (loss of data)

treat like missing value

discretization

Capping

outliers

Z Score treatment

IQR based filtering

Percentile

Winsorization

→ It depends upon problem statement. For example in case of anomaly detection, outliers are useful.

How to detect outliers

if data is normally distributed

OneNote for Windows 10
nitishsingh24@gmail.com

How to detect Outliers?
Thursday, April 29, 2021 3:27 PM

1. Normal Distribution

$\mu + 3\sigma \leftarrow \}$ outliers
 $\mu - 3\sigma \leftarrow \}$ outliers



OneNote for Windows 10
nitishsingh24@gmail.com

2. Skewed Distribution

$\text{min} = Q_1 - 1.5 \times IQR$
 $Q_3 + 1.5 \times IQR$



If data is having skewed distribution

OneNote for Windows 10
nitishsingh24@gmail.com

3. Other Distributions



for other distribution