

Back

Outliers Quiz

Graded Quiz • 30 min

Hide menu

Outliers

Video: Outliers

16 min

Reading: Outliers Demo

1h

Reading: Outliers Case Study - CC Fraud Detection

1h

Quiz: Outliers Quiz

Submitted

Discussion Prompt: Outliers Exploration Exercise

2h

## Outliers Quiz

### Review Learning Objectives

Submit your assignment

Due

Feb 25, 11:59 PM IST

Receive grade

To Pass

60% or higher

Like

Dislike

Report an issue

Congratulations! You passed!

Grade received

100%

Latest Submission Grade

100%

To pass

60% or higher

Go to next item

1.

What are outliers in data analysis?

1 / 1 point

- ☐ Outliers are the most frequent data points in a dataset.
- ☐ Outliers are the data points that lie within the normal range of the data distribution.
- ☒ Outliers are the data points that do not conform to the majority of the data in the dataset.
- ☐ Outliers are the data points that have the highest correlation with the target variable.

☒ Correct

Correct! Outliers are data points that do not follow the majority of the data in the dataset.

Try again

Your grade

100%

View Feedback

We keep your highest score

2.

Which of the following methods is used to detect outliers based on the standard deviation from the mean?

1 / 1 point

- ☐ IQR (Interquartile Range)
- ☐ DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- ☒ Z-score
- ☐ LOF (Local Outlier Factor)

☒ Correct

Correct! Z-score is used to detect outliers based on the standard deviation from the mean.

3.

What is the Z-score used for in outlier detection?

1 / 1 point

- ☐ To identify the outliers that have a distance greater than a specified threshold.
- ☒ To standardize data and calculate how many standard deviations a data point deviates from the mean.
- ☐ To partition the data into clusters based on density.
- ☐ To calculate the correlation between two variables in the dataset.

☒ Correct

Correct! The Z-score calculates how many standard deviations a data point deviates from the mean, helping identify outliers.

4.

Which of the following is a drawback of using Z-score for outlier detection?

1 / 1 point

- ☐ It cannot handle high-dimensional datasets.
- ☐ It requires a large amount of labeled training data.
- ☒ It assumes that the data is normally distributed.
- ☐ It is computationally expensive for large datasets.

☒ Correct

Correct! Z-score assumes that the data follows a normal distribution, which may not always be the case.

5.

What is the IQR (Interquartile Range) used for in outlier detection?

1 / 1 point

- ☒ To identify the outliers that have a distance greater than a specified threshold.
- ☐ To standardize data and calculate how many standard deviations a data point deviates from the mean.
- ☐ To partition the data into clusters based on density.
- ☐ To calculate the correlation between two variables in the dataset.

☒ Correct

This option is correct! The IQR is used to identify outliers based on a specified threshold of the interquartile range.

6.

What is the One-Class SVM (Support Vector Machine) used for in outlier detection?

1 / 1 point

- ☐ To identify the outliers that have a distance greater than a specified threshold.
- ☒ To classify data points as either normal or outliers based on a training set of only normal data.
- ☐ To partition the data into clusters based on density.
- ☐ To calculate the correlation between two variables in the dataset.

☒ Correct

Correct! The One-Class SVM is used for novelty detection by classifying data points as normal or outliers based on a training set of only normal data.

7.

What is the Isolation Forest used for in outlier detection?

1 / 1 point

- ☐ To identify the outliers that have a distance greater than a specified threshold.
- ☐ To classify data points as either normal or outliers based on a training set of only normal data.
- ☐ To partition the data into clusters based on density.
- ☒ To isolate outliers by randomly selecting features and splitting data points into partitions.

☒ Correct

Correct! The Isolation Forest isolates outliers by randomly selecting features and splitting data points into partitions, making it efficient for detecting outliers.

8.

Which of the following is a limitation of using Isolation Forest for outlier detection?

1 / 1 point

- ☐ It cannot handle high-dimensional datasets.
- ☐ It requires a large amount of labeled training data.
- ☐ It assumes that the data is normally distributed.
- ☒ It may have difficulty detecting outliers in datasets with clusters of similar data points.

☒ Correct

Correct! Isolation Forest may struggle to detect outliers in datasets with clusters of similar data points, as it relies on random partitioning.

9.

Which outlier detection method is based on the density of data points in the feature space?

1 / 1 point

- ☐ Z-score
- ☐ IQR (Interquartile Range)
- ☒ DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- ☐ LOF (Local Outlier Factor)

☒ Correct

Correct! DBSCAN is based on the density of data points in the feature space for outlier detection.

10.

Which of the following is a limitation of using DBSCAN for outlier detection?

1 / 1 point

- ☐ It cannot handle high-dimensional datasets.
- ☐ It requires a large amount of labeled training data.
- ☐ It assumes that the data is normally distributed.
- ☒ It may not work well with datasets of varying density or irregular shapes.

☒ Correct

Correct! DBSCAN may not perform well with datasets of varying density or irregular shapes.

11.

Which outlier detection method is based on the concept of measuring the local deviation of a data point with respect to its neighbors?

1 / 1 point

- ☐ Z-score
- ☐ IQR (Interquartile Range)
- ☐ DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- ☒ LOF (Local Outlier Factor)

☒ Correct

Correct! LOF is based on measuring the local deviation of a data point with respect to its neighbors for outlier detection.

12.

Which of the following is a drawback of using LOF for outlier detection?

1 / 1 point

- ☐ It cannot handle high-dimensional datasets.
- ☐ It requires a large amount of labeled training data.
- ☐ It assumes that the data is normally distributed.
- ☒ It may have difficulty detecting outliers in datasets with clusters of similar data points.

☒ Correct

Correct! LOF may struggle to detect outliers in datasets with clusters of similar data points, as it measures local deviations relative to neighbors.

