

# Anomaly Detection Quiz

<b>Due</b> No due date	<b>Points</b> 3	<b>Questions</b> 3
<b>Available</b> after May 15 at 11:32		<b>Time Limit</b> None
<b>Allowed Attempts</b> Unlimited		

Take the Quiz Again

## Attempt History

	Attempt	Time	Score
LATEST	<a href="#">Attempt 1</a>	less than 1 minute	0 out of 3

Submitted Jun 16 at 18:17

Unanswered

Question 10 / 1 pts

Which one of the following strategies in anomaly detection does **not** calculate the **distance** between data points to determine an anomalous vs. a normal instance?

☐ Density-based approaches

Correct Answer

☐ Statistical approaches

☐ Proximity-based approaches

☐ Clustering-based approaches

In Statistical methods, we assume that the normal data follows some statistical model and we report instances as anomalous if they occur in low-probability regions of the distribution. However, in proximity-based, Density-based, and Clustering-based approaches, we always calculate the distance between data points to determine whether an instance is far away (proximity-based), in low density regions (density-based), or in a small cluster and far from others (clustering-based).

Unanswered

## Question 2

0 / 1 pts

In Relative density outlier score (Local Outlier Factor, LOF), why do we consider the distance that an instance's nearest neighbour has to their nearest neighbours? In other words, why do we have the following denominator for LOF score?

$$relative\ density(x, k) = \frac{density(x, k)}{\frac{1}{k} \sum_{y \in N(x, k)} density(y, k)}$$

Correct Answer

- ☐ to detect anomalies even with various density regions
- ☐ to detect Collective anomalies
- ☐ to remove the sensitivity to the number of nearest neighbours

The reason for considering the distances of nearest neighbours to their corresponding nearest neighbours is to account for regions in the data that have variable densities. For instance in a region, where the nearest neighbours themselves are compactly clustered, an instance should be penalised much more than when the nearest neighbours themselves are in a low density cluster. This method does not remove the sensitivity to the number of nearest neighbours or the radius within which nearest neighbours are selected.

Unanswered

### Question 3

0 / 1 pts

How can we handle anomaly detection with respect to variable density clusters in an unsupervised algorithm like k-means?

- ☐ Perform a hill climbing search with multiple initializations
- ☐ Remove small and far clusters
- ☐ Use relative distance instead of absolute distance
- ☐ Remove objects which most improve the clustering objective function

Correct Answer

To remove the sensitivity of distance with respect to variable density regions, we can always use relative distance. All of the other techniques can be used for the detection of best clusters for the normal data, they help with an algorithm such as k-means not to get stuck in local optima.