

OutlierDetection: An R package for scoring and visualization of

Outliers

Vinay Tiwari

Department of Statistics, Savitribai Phule Pune University

Abstract

Outliers are data points that differ significantly from other observations. Outliers have special importance in statistical analysis as either they sometimes hold key properties related to the data or their presence hampers the entire analysis and the analyst ends up getting incorrect results. Hence, Outlier Detection is an important step in any statistical analysis exercise. Outlier Detection can be done in a Supervised, Unsupervised and Semi Supervised manner.

This study focuses on unsupervised learning methods which tag an observation as outlier based on its statistical closeness to the rest of the data points. These statistical closeness definitions give rise to variety of methods available in the literature.

We have developed an R package which has the following characteristics-

1. It does outlier detection using variety of methods
2. It provides outlier score/ probability using non-parametric bootstrapping
3. It has 2D and 3D visualizations for the outliers

Keywords: Data science, Visualization, Outlier detection, Anomaly detection,

Unsupervised learning