PyNomaly: Anomaly detection using Local Outlier Probabilities (LoOP).

Valentino Constantinou[1]

1 NASA Jet Propulsion Laboratory

Summary

`PyNomaly` is a Python 3 implementation of LoOP (Local Outlier Probabilities). LoOP is a local density based outlier detection method by Kriegel, Kröger, Schubert, and Zimek which provides outlier scores in the range of [0,1] that are directly interpretable as the probability of a sample being an outlier. `PyNomaly` also implements a modified approach to LoOP introduced by Hamlet et. al., which may be used for applications involving streaming data or where rapid calculations may be necessary.

The outlier score of each sample is called the Local Outlier Probability. It measures the local deviation of density of a given sample with respect to its neighbors as Local Outlier Factor (LOF), but provides normalized outlier scores in the range [0,1]. These outlier scores are directly interpretable as a probability of an object being an outlier. Since Local Outlier Probabilities provides scores in the range [0,1], practitioners are free to interpret the results according to the application.

Like LOF, it is local in that the anomaly score depends on how isolated the sample is with respect to the surrounding neighborhood. Locality is given by k-nearest neighbors, whose distance is used to estimate the local density. By comparing the local density of a sample to the local densities of its neighbors, one can identify samples that lie in regions of lower density compared to their neighbors and thus identify samples that may be outliers according to their Local Outlier Probability.

`PyNomaly` includes an optional *cluster_labels* parameter. This is useful in cases where regions of varying density occur within the same set of data. When using *cluster_labels*, the Local Outlier Probability of a sample is calculated with respect to its cluster assignment.

Research

PyNomaly is currently being used in the following research:

- Y. Zhao and M.K. Hryniewicki, "XGBOD: Improving Supervised Outlier Detection with Unsupervised Representation Learning," International Joint Conference on Neural Networks (IJCNN), IEEE, 2018.

References