Notes on Anomaly Detection

Sameer Kesava PhD

April 26, 2020

Contents

Supervised Learning Algorithms

1.1 Univariate Data

- Boxplot
- Grubbs test

1.2 Multivariate Data

1.3 Random Cut Forest

From Amazon SageMaker

1.4 XGBoost

- Gradient boosting method
- Absolute loss and Huber loss more robust to outliers.
- Hyperparameters
    1. Max_depth
    2. Colsample_bytree
    3. Eta
    4. train-test split: 60-40/70-30/80-20.

Unsupervised Learning Algorithms

2.1 MeanShift Clustering

2.2 DBSCAN

Improving the Accuracy

3.1 Hyperparameter Tuning

- Hyperparameter optimization based on Gaussian Process Regression and Bayesian Optimization
- keras tuner in keras
- GridSearchCV or RandomSearchCV in scikit-learn
- RandomSearch can be used as the baseline against which optimization algorithms can be evaluated.

Bibliography

[1] Pankaj Malhotra et al., Long Short Term Memory Networks for Anomaly Detection in Time Series, 2015.