Anomaly Detection Challenges 2016/17

Chair of IT Security, TU München

Homework 1

As you have seen in class, there are many anomaly detection approaches developed by adapting existing machine learning methods or inventing new ones. Your task for this homework is to test the unsupervised anomaly detection approaches (ignore labels in training) using the dataset nr. 514 from the following web page:

http://homepage.tudelft.nl/n9d04/occ/index.html.

The dataset is contained in the MAT file, which you can load using the *loadmat* function of *scipy* library. This dataset is related to distinguishing presence from absence of Arrhytmia using various diagnostic features. Use the following methods to discover anomalies:

1) Kernel Density Estimation with Gaussian kernel (use scikit-learn or try to implement it yourself)

2) One-Class SVM (use scikit-learn or try to implement it yourself)

3) Local Outlier Factor (try to implement it yourself)

Test the performance of these methods and compare their performance. Is there a significant difference? Comment on this difference. How about the difference in running time?

Send solutions (use jupyter or ipython notebook) to kolosnjaji@sec.in.tum.de and describe in your e-mail or in the notebook shortly how you did it and comment on your results. Maybe also add matplotlib plots to make it more convincing.

Have fun!