Anomaly Detection Challenges 2015/16

Chair of IT Security, TU München

Homework 3

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 using the dataset nr. 514 from the following web page:

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

The data set 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, or your improved implementation of it from homework 1.
2. One-Class SVM (you can use the implementation from scikit-learn)
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.

Send solutions (code, if possible 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 add matplotlib plots to make it more convincing.

Have fun!