

## Homework 4

There are many anomaly detection problems where the dataset is partially labeled. Your task for the next homework is to test the semi-supervised anomaly detection approaches, again 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.

Try to randomly select labels so that labeled examples constitute:

- a) 10%
- b) 30%
- c) 50%
- d) 100%

of the training set.

Divide the dataset randomly in half, 50% for the training set and 50% for the test set.

a) Use label propagation from scikit-learn to test the change in accuracy w.r.t. the percent of labelled data. Repeat the test 100 times and show average results.

b) (Optional) Implement a semi-supervised version of KDE based on the paper mentioned in class and test it using the generated training and test sets. You can use basic KDE, or your implementation from homework 1.

Have fun!