Technische Hochschule Ostwestfalen-Lippe
Fachbereich Elektrotechnik und Technische Informatik
Prof. Dr. Helene Dörksen

Exercises and Labs 11
for Lecture "Authentication „ (M.Sc.)

Lab 11.1 Consider dataset *breast_cancer.dat* and corresponding script *breast_cancer_load.m*. In this script you will also find an example for the *convex hull* calculation.

You might also treat samples *xyz.dat* provided previously and corresponding scripts *xyz_load.m* as well as other datasets of your choice. Take into account, that most interesting are such samples with large number of objects.

Consider initial classifier and refinement within possible combinations:

initial by SVM & refinement by SVM;

initial by LDA & refinement by SVM;

initial by SVM & refinement by LDA.

Apply *ComRef* for fusion of summands and dimensionality reductions to 2D feature spaces.

a) Calculate refinements of initial classifier in 2D feature spaces. For one of the 2D feature spaces visualise objects, convex hulls of classes, initial classifier and its refinement.

b) Analyse the accuracies of initial classifier, its corresponding refinements and refinements calculated on vertices of convex hulls.

c) Compute speed-ups for refinements calculated on vertices of convex hulls.

Remark: You will find useful information for these labs in the following publication:

Helene Dörksen, Uwe Mönks and Volker Lohweg: "Fast Classification in Industrial Big Data Environments," in 19th IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA 2014), 2014.

Lab 11.2 Classifier Design (advanced self-study) Consider dataset *MAGIC_gamma.dat* and corresponding script *MAGIC_gamma_load.m*.

Execution time for *SVM* calculation is extremely long. The classification accuracy of initial *LDA* is about 79.48% and seems to outperform *SVM*.

Try to find a linear classifier (or combination of linear classifiers), which outperforms initial *LDA*. Idea: split classes in parts (have a look on the histograms of features).

Lab 11.3 Classifier Design (advanced self-study) Consider dataset *Sensorless_drive_diagnosis.txt* and corresponding script *Motor_Drive_Diagnosis_load.m*.

This sample is collected for condition monitoring of running motors in industrial environments. It contains 11 classes, which are *anomaly* (class labels $1-10$) and *intact* (class label 11).

In the document $Initial\_accuracies\_for\_pairwise\_classification.dat$ you will find information about accuracies of initial *SVM* and *LDA* regarding all possible pairwise classification tasks. As you can seen in this document, there are such pairs, where initial classifiers fail. E.g., for classification anomaly classes 1 and 6 the initial accuracies are 90.36% for *SVM* and 95.90% for *LDA*. However, in the real application scenarios it is important to recognise that exactly kind of anomaly appears and, thus, to fix the shortcoming correctly.

Your task is to design an accurate linear classifier (or combination of linear classifiers), which is as accurate as possible and has preferably small number of features, to solve these problems.