

## Exercises and Labs 6 for Lecture “Authentication „ (M.Sc.)

**Exercise 6.1** PCA (Principal Components Analysis) and LDA (Linear Discriminant Analysis) have different purposes. The former operates on the total (unlabeled) dataset to find the directions that contain the maximum variance; the latter operates on labeled data to find the directions which are best at distinguishing the labeled classes. In general the results (principal components and canonicals, respectively) will be quite different. However, for special examples of data the principal components and canonicals could be in the same directions.

- (i) Draw an example of two-class, two-dimensional data such that PCA and LDA find the same directions.
- (ii) Draw an example of two-class, twodimensional data such that PCA and LDA find totally different directions.

**Lab 6.2** Consider datasets *banknote.dat*, *hepatitis.dat* and *seeds.dat* from different authentication applications.

Corresponding scripts are  
*banknote\_load.m*, *hepatitis\_load.m* and *seeds\_load.m*

More details about datasets are found on  
<https://archive.ics.uci.edu/ml/index.php>

Helpful Matlab functions are found in *banknote\_load.m*. Additional Matlab functions are provided.

For given datasets:

- a) Display histograms of features in original form as well as within normalization and softmax scaling (as given in Lecture 6).
- b) What does the *balance* of dataset mean? Are datasets well-balanced?
- c) Calculate FDR for single features as well as for arbitrary feature combinations.
- d) Classify the datasets within *LDA* and *SVM* approaches. Compare their accuracies. Calculate FDRs for *LDA* and *SVM* solutions.

**Lab 6.3 advanced self-study** Consider dataset *AmazonCommerce.txt* for the investigation of Amazon customer/buying behavior.

Corresponding script is *AmazonCommerce\_load.m*

More details about this dataset are found on  
<https://archive.ics.uci.edu/ml/index.php>

**Remark:** it might take longer time during loading the dataset as well during the executions of results.

- a) Display histograms of features in original form as well as within normalization and softmax scaling.
- b) Classify the dataset within *LDA* and *SVM* approaches. Did it work well?
- c) Think about possible techniques to solve the classification task.