Exercises for
Pattern Recognition
Peter Fischer, Shiyang Hu
Assignment 5, 18./21.11.2014

SOLUTION

General Information:

Exercises (1 SWS):	Tue 12:15 – 13:45 (0.154-115) and Fri 08:15 – 09:45 (0.151-115)
Certificate:	Oral exam at the end of the semester
Contact:	peter.fischer@fau.de, shiyang.hu@fau.de

Fisher Transform and Norms

Exercise 1 In this exercise, we derive the Fisher transform for dimensionality reduction of feature vectors. The Fisher transform is a formulation of the LDA. In the training step, the class label for each pattern is known. The optimal projection axis $\boldsymbol{a}^*$ for the Fisher transform is calculated according to the Rayleigh ratio:

$$\boldsymbol{a}^* = \operatorname*{argmax}_{\boldsymbol{a}} \frac{\boldsymbol{a}^T \boldsymbol{\Sigma}_{\mathtt{inter}} \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma}_{\mathtt{intra}} \boldsymbol{a}}$$

The definitions for $\boldsymbol{\Sigma}_{\mathtt{inter}}$ and $\boldsymbol{\Sigma}_{\mathtt{intra}}$ can be found in the lecture slides. We are only interested in the 2 class problem.

(a) Describe the different quantities in the Rayleigh ratio.

$\boldsymbol{\Sigma}_{\mathtt{inter}}$: covariance matrix of transformed mean vectors, measure of how far the class means are apart
$\boldsymbol{\Sigma}_{\mathtt{intra}}$: mean of covariance matrices of all the classes, measure of how far each class is scattered
Rayleigh ratio: find a transformation where the mean vectors of the two classes are far apart and each class is close to its mean. Therefore, the overlap between the classes is minimized.

(b) Reformulate the given (unconstrained) optimization problem to a constrained optimization problem using Lagrange multipliers.

$$\boldsymbol{a}^* = \operatorname*{argmax}_{\boldsymbol{a}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{\mathtt{inter}} \boldsymbol{a} - \lambda \boldsymbol{a}^T \boldsymbol{\Sigma}_{\mathtt{intra}} \boldsymbol{a}$$

(c) Solve the constrained problem to determine $\boldsymbol{a}^*$.
Hint: derivatives w.r.t. $\boldsymbol{a}$ must be 0. Use the Matrix Cookbook.
Use the identity (for symmetric $\boldsymbol{\Sigma}$):

$$\frac{\partial \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}}{\partial \boldsymbol{a}} = 2\boldsymbol{\Sigma}\boldsymbol{a}$$

The derivative gives:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{a}}\left(\boldsymbol{a}^T \boldsymbol{\Sigma}_{\texttt{inter}} \boldsymbol{a} - \lambda \boldsymbol{a}^T \boldsymbol{\Sigma}_{\texttt{intra}} \boldsymbol{a}\right) &= 0 \\
2\boldsymbol{\Sigma}_{\texttt{inter}} \boldsymbol{a} - 2\lambda \boldsymbol{\Sigma}_{\texttt{intra}} \boldsymbol{a} &= 0 \\
\boldsymbol{\Sigma}_{\texttt{inter}} \boldsymbol{a} &= \lambda \boldsymbol{\Sigma}_{\texttt{intra}} \boldsymbol{a} \\
\boldsymbol{\Sigma}_{\texttt{intra}}^{-1} \boldsymbol{\Sigma}_{\texttt{inter}} \boldsymbol{a} &= \lambda \boldsymbol{a}
\end{aligned}$$

This is an eigenvalue-problem and can be solved with standard methods. In the special case of 2 classes, because $\boldsymbol{\Sigma}_{\texttt{inter}}$ is of rank 1 and thus $\boldsymbol{\Sigma}_{\texttt{inter}} \boldsymbol{a} \propto \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, the solution can directly be written as:

$$\boldsymbol{a}^* = \boldsymbol{\Sigma}_{\texttt{intra}}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

Exercise 2 Implement a classification algorithm which classifies a new feature only by calculating the distance to the class prototypes (i.e., the mean of a class). Use the option to pass parameters to the classification function, and implement three approaches to calculate the distance:

(a) the L1-Norm

(b) the L2-Norm (= euclidean distance)

(c) the Mahalanobis distance, which incorporates the covariance matrices.

Solution implemented in NormNeighbor.m