Prof. Dr. Vladimir Shikhman
Professur für Wirtschaftsmathematik
Technische Universität Chemnitz

Übungsleiter: David Müller
david.mueller@mathematik.tu-chemnitz.de

Mathematical Foundation of Big Data Analytics (SS 2019)
Classification I

Ex. 1 Given a classification problem with two classes C_1 and C_2 . The p-dimensional Gaussian density is

$$f_\ell(x) = \frac{1}{(2\pi)^{p/2} \cdot \sqrt{\det(\Sigma_\ell)}} \cdot \exp(-\frac{1}{2}(x-\mu_\ell)^T \cdot \Sigma_\ell^{-1} \cdot (x-\mu_\ell)^T).$$

Recall that in a general classification with m-classes a new observation of features $x \in \mathbb{R}^p$ belongs to class $C_{\hat{\ell}}$ if

$$f_{\hat{\ell}}(x) = \max_{\ell=1,\dots,m} f_\ell(x).$$

a) Show that for case of two classes under the assumption of constant covariance matrix, i. e. $\Sigma_1 = \Sigma_2 = \Sigma$, the Gaussian discriminant becomes a linear classifier.

b) Show that the solution of a) is equivalent to the Fisher discriminant.

Ex. 2 Consider the following 2-dimensional dataset, which is divided into two classes

$$C_1 = \{(4,2)^T, (2,4)^T, (2,3)^T, (3,6)^T, (4,4)^T\}$$
$$C_2 = \{(9,10)^T, (6,8)^T, (9,5)^T, (8,7)^T, (10,8)^T\}$$

Calculate the linear discriminant for the dataset.
Hint: For two classes the matrix

$$B = \sum_{\ell=1}^{2} n_\ell(\bar{x}_\ell - \bar{x}) \cdot (\bar{x}_\ell - \bar{x})^T$$

can (up to a constant) be written as

$$\tilde{B} = (\bar{x}_1 - \bar{x}_2) \cdot (\bar{x}_1 - \bar{x}_2)^T.$$

Ex. 3 The Soft-Margin-SVM problem is given by

$$\begin{aligned} \min_{w,b,\epsilon} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \epsilon_i \\ s.t \quad & y_i(w^T x_i + b) \geq 1 - \epsilon_i, \quad i = 1,\dots,m \\ & \epsilon_i \geq 0, \quad i = 1,\dots,m \end{aligned}$$

a) Derive the dual problem.

b) Derive the possible values for the dual variables.