

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

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

Mathematische Grundlagen von Big Data Analytics (SS 2018)
Übung 7: Classification I

1) Gegeben sei folgender zweidimensionaler Datensatz, der in zwei Klassen aufgeteilt ist:

$$C_1 = \{(4, 1), (2, 4), (2, 3), (3, 6), (4, 4)\}$$
$$C_2 = \{(9, 10), (6, 8), (9, 5), (8, 7), (10, 8)\}$$

Berechnen Sie die lineare Diskriminante für den Datensatz.

2) Betrachten Sie das sogenannte Soft Margin SVM Problem. Im Vergleich zum Hard SVM können hier einige Punkte die Constraints verletzen:

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

a) Bestimmen Sie das duale Problem.

b) Erläutern Sie anhand einer Fallunterscheidung, welche Werte die Dualvariablen annehmen können.

3) Bestimmen Sie für zweidimensionale Merkmale die Feature Map für

a) den euklidischen Kernel.

b) den Gauss Kernel.