Prof. Dr. Katharina Morik,
Prof. Dr. Claus Weihs,
Dr. Wouter Duivesteijn,
M.Sc. Sarah Schnackenberg,
B.Sc. Melanie Dagge

Dortmund, 28.05.14
Abgabe: bis Do, 04.06.2015,
10 Uhr, an
`wouter.duivesteijn@tu-dortmund.de`
und/oder in den Briefkasten "Duivesteijn"
im OH12, R4.005

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2015
Blatt 7

Wiederholung Die *Support Vector Machine (SVM)* oder auch *Stützvektormethode* ist ein komplexes Lernverfahren, das nahe an der Linearen Regression liegt. Im Folgenden sollten Sie sich einige Gedanken zum Prinzip der Klassifikation mit dem SVM-Verfahren machen:

1. Wozu wird die SVM verwendet und was ist das Ergebnis des Lernschrittes?
2. Erklären Sie den Zusammenhang der separierenden Hyperebene mit dem *maximum margin problem*!
3. Wie wird das *maximum margin problem* gelöst? Was ist das *duale Problem*?
4. Was drückt die Bedingung
$$\sum_{i=1}^{n} \alpha_i y_i = 0$$
in Bezug auf die Berücksichtigung der Stützvektoren aus?

Hinweis: Von den in der Vorlesung vorgestellten Arbeiten finden Sie (fast) alle über `https:/scholar.google.de` aus dem Uni-Netz. Unterstützend zur Vorlesung können folgende Papiere gelesen werden:

- *"A Tutorial on Support Vector Machines for Pattern Recognition"*, Christopher Burges in: "Data Mining and Knowledge Discovery", 1998, pp. 121-167
- *"Fast Training of Support Vector Machines using Sequential Minimal Optimization"*, John C. Platt in: "Advances in Kernel-Methods", 1999, MIT Press, pp. 185-208
- *"Advances in Kernel-Methods"*, Schölkopf, Burges, Smola (eds), 1999, MIT Press
- *"Theorie der Zeichenerkennung"*, W.N. Vapnik, A. Tscherwonenkis, 1979, Akademie Vg. [UB-Dortmund: Signatur M20947]

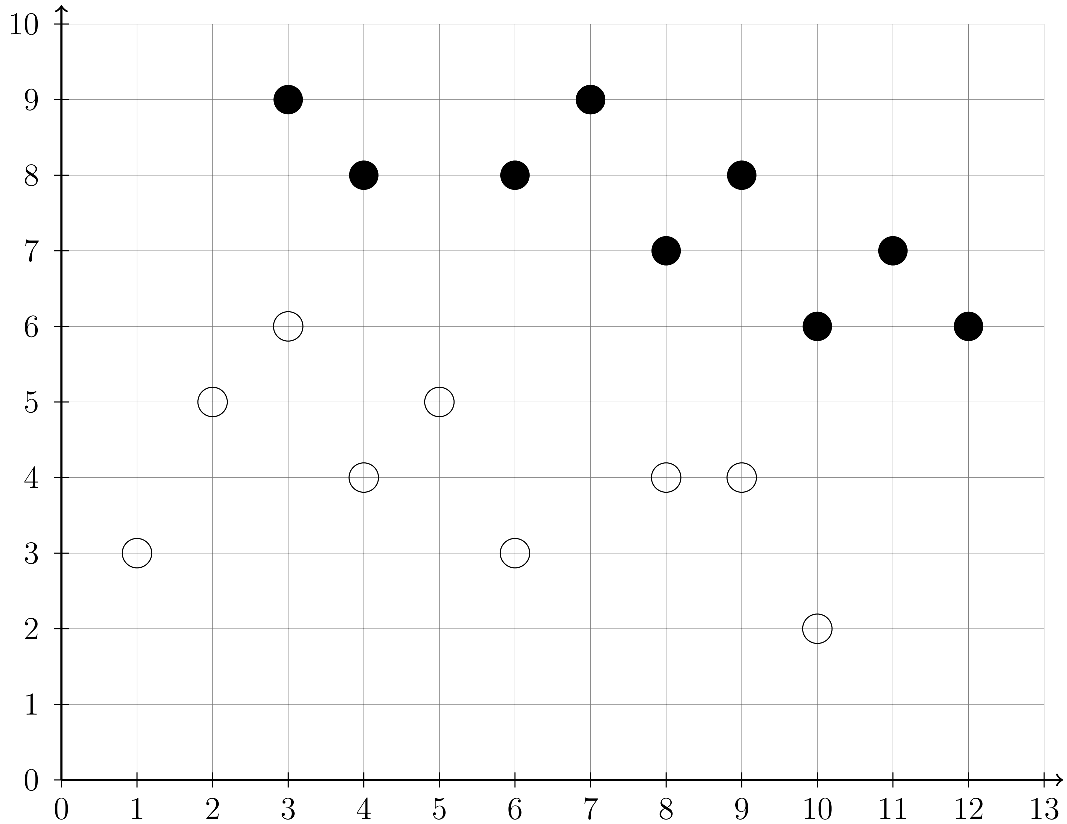

Abbildung 1: Datenpunkte im $\mathbb{R}^2$

Aufgabe 7.1 (10 Punkte)
In dieser Aufgabe sollen Sie den naiven Vorläufer des SVM-Algorithmus "von Hand" ausprobieren. Gegeben ist die folgende Menge D von Datenpunkten (vgl. Abbildung 1):

$$\begin{aligned} D \ = \{ \quad & (1,3,-1),(2,5,-1),(3,6,-1),(4,4,-1), \\ & (5,2,-1),(6,3,-1),(8,4,-1),(9,4,-1), \\ & (7,3,-1),(5,5,-1),(6,9,+1),(11,6,+1), \\ & (3,9,+1),(4,8,+1),(6,8,+1),(7,9,+1), \\ & (8,7,+1),(9,8,+1),(10,6,+1),(11,7,+1) \ \} \end{aligned}$$

Dabei bezeichnen die Komponenten jedes Tripels (x_1, x_2, y) aus D die erste und zweite Koordinate des Punktes sowie die zugehörige Klasse $y \in \{-1, 1\}$.

1. (3 Punkte) Bestimmen Sie für $j \in \{-1, +1\}$ jeweils die Mittelpunkte $\vec{c}_+$ und $\vec{c}_-$ der Menge
$$C_j = \{(x_1, x_2, y) \in D \mid y = j\}$$
2. (4 Punkte) Bestimmen Sie $\vec{\beta}$ und den Mittelpunkt $\vec{\alpha}$.
3. (3 Punkte) Zu welchen Klassen werden anhand dieses einfachen Verfahrens die folgenden Punkte zugeordnet?
$$(4,6),(7,6),(12,4),(-1,8),(-4,11)$$