

Data Mining Block 3

A1.1

$$C_- \quad x_{C_-} = \frac{1+2+3+4+5+6+8+9+7+5}{10} = \frac{50}{10} = 5$$

$$y_{C_-} = \frac{8+5+6+4+2+3+4+4+3+5}{10} = \frac{39}{10} = 3,9$$

$$C_+ \quad x_{C_+} = \frac{6+11+3+4+6+7+8+9+10+11}{10} = \frac{75}{10} = 7,5$$

$$y_{C_+} = \frac{9+6+9+8+8+9+7+8+6+7}{10} = \frac{77}{10} = 7,7$$

$$C_- (5/3,9) \quad C_+ (7,5/7,7)$$

A1.2

$$\vec{w} = C_+ - C_- \Rightarrow \begin{aligned} w_x &= 7,5 - 5 = 2,5 \\ w_y &= 7,7 - 3,9 = 3,8 \end{aligned} \quad w = (2,5/3,8)$$

$$C = \frac{C_+ + C_-}{2}$$

$$c_x = \frac{5 + 7,5}{2} = 6,25$$

$$c_y = \frac{3,9 + 7,7}{2} = 5,8$$

A1.3

$$x'_1 = x_1 - C \quad \begin{pmatrix} 4 - 6,25 \\ 6 - 5,8 \end{pmatrix} \quad x'_1 = \begin{pmatrix} -2,25 \\ 0,2 \end{pmatrix}$$

$$\langle w, x'_1 \rangle = -2,25 \cdot 2,5 + 0,2 \cdot 3,8 = -4,865$$

$$x'_2 = \begin{pmatrix} 7 - 6,25 \\ 6 - 5,8 \end{pmatrix} \quad x'_2 = \begin{pmatrix} 0,75 \\ 0,2 \end{pmatrix}$$

$$\langle w, x'_2 \rangle = 0,75 \cdot 2,5 + 0,2 \cdot 3,8 = 2,635$$

$$x'_3 = \begin{pmatrix} 12 - 6,25 \\ 4 - 0,2 \end{pmatrix} \quad \begin{pmatrix} 5,75 \\ 3,8 \end{pmatrix} = x'_3$$

$$\langle w, x'_3 \rangle = 5,75 \cdot 2,5 + 3,8 \cdot 3,8 = 7,535$$

$$x_4' = \begin{pmatrix} -1 - 6,25 \\ 8 - 5,8 \end{pmatrix} = \begin{pmatrix} -7,25 \\ 2,2 \end{pmatrix}$$

$$\langle W, x_4' \rangle = 2,5 \cdot -7,25 + 3,8 \cdot 2,2 = -18,125 + 8,36 = -9,765$$

$$x_8' = \begin{pmatrix} -4 - 6,25 \\ 11 - 5,8 \end{pmatrix} = \begin{pmatrix} -10,25 \\ 5,3 \end{pmatrix}$$

$$\langle W, x_8' \rangle = 2,5 \cdot -10,25 + 3,8 \cdot 5,3 = -25,625 + 20,14 = -5,485$$

Übungsblatt 3

SVM 25 Punkte

Data Mining

Wintersemester 2016/17

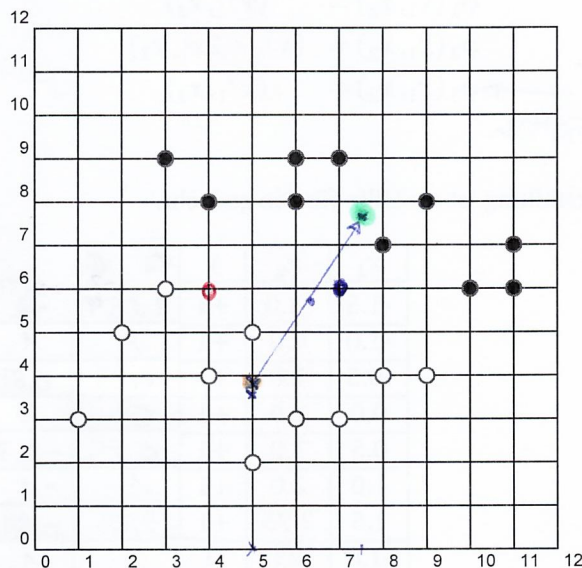
Abgabe: 02.12.2016 9:45 Uhr

Aufgabe 1 SVM-Vorläufer 7 Punkte

In dieser Aufgabe sollen Sie den naiven Vorläufer des SVM-Algorithmus „von Hand“ ausprobieren. Gegeben ist die folgende Menge D von Datenpunkten:

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

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



$$50/10 = 5$$

$$36/10 = 3.6$$

$$C_-(5/3, 9)$$

$$C_+(7, 5/7, 7)$$

$$C(6, 25/5, 8)$$

Aufgabe 1.1 4 Punkte

Bestimmen Sie für $j \in \{-1, +1\}$ jeweils die Mittelpunkte \vec{c}_+ und \vec{c}_- der Menge

$$C_j = \{(x_1, x_2, y) \in D | y = j\}$$

Aufgabe 1.2 2 Punkte

Bestimmen Sie \vec{w} und den Mittelpunkt \vec{c} .

Aufgabe 1.3 1 Punkt

Zu welchen Klassen werden anhand dieses einfachen Verfahrens die folgenden Punkte zugeordnet?

$$(4, 6), (7, 6), (12, 4), (-1, 8), (-4, 11)$$

Aufgabe 2 Kernel-Funktionen 5 Punkte

Eine Stärke der SVM ist die Verwendung von Kernel-Funktionen, die eine implizite Transformation ϕ der Daten in einen anderen Raum ermöglichen, so dass ursprünglich linear nicht-trennbare Daten in diesem neuen Raum schließlich trennbar sind. In dieser Aufgabe geht es darum, den Effekt der ϕ -Transformation der Datenpunkten zu untersuchen. Gegeben sind die Datenpunkte aus der Tabelle.

Transformieren Sie die Daten mit den nachfolgenden Funktionen ϕ_i und geben Sie die transformierte Tabelle sowie eine grafische Darstellung der neuen Punkte an. Zeichnen Sie die transformierten Punkte in die Koordinatensysteme auf dem Aufgabenblatt ein.

Welche Funktion ermöglicht eine Trennung der Daten?

Trennung der Daten ist möglich →

$$\begin{aligned}\phi_1(x_1, x_2) &= (x_1^2, x_2) \\ \phi_2(x_1, x_2) &= (x_1^3 - 2x_1, x_2) \\ \phi_3(x_1, x_2) &= (x_1^3, x_2)\end{aligned}$$

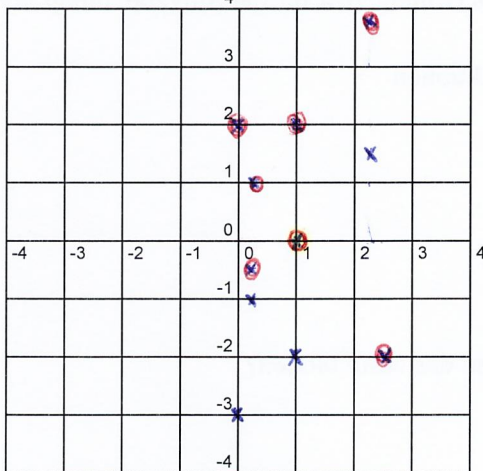
Hinweis: Die graphische Darstellung ist als Hilfe für Sie gedacht.

x_1	x_2	y	ϕ_1	ϕ_2	ϕ_3
-1.5	-2.0	+1	2.25	-3	-3.375
-1.0	0.0	+1	1	-2	-1
-0.5	1.0	+1	0.25	-1	-0.125
0.0	2.0	+1	0	0	0
0.5	1.0	+1	0.25	-1	0.125
1.0	2.0	+1	1	0	1
1.5	3.75	+1	2.25	3.375	3.375
-1.0	-2.0	-1	1	-2	-1
-0.5	-1.0	-1	0.25	-1	-0.125
0.0	-3.0	-1	0	0	0
0.5	-0.5	-1	0.25	-1	0.125
1.0	-2.0	-1	1	0	1
1.5	1.5	-1	2.25	3.375	3.375

Tabelle 1: Instanzen

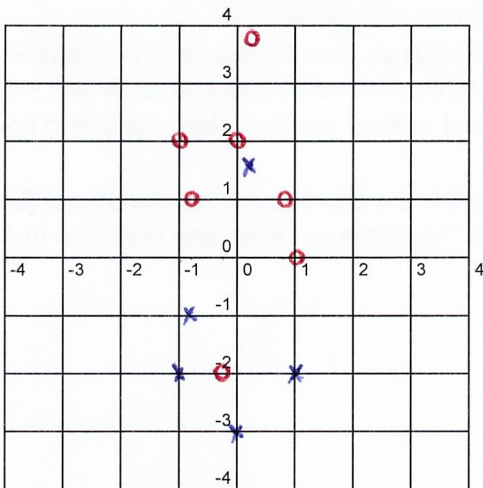
Φ_1

$$\begin{aligned} O &= +1 \\ X &= -1 \end{aligned}$$



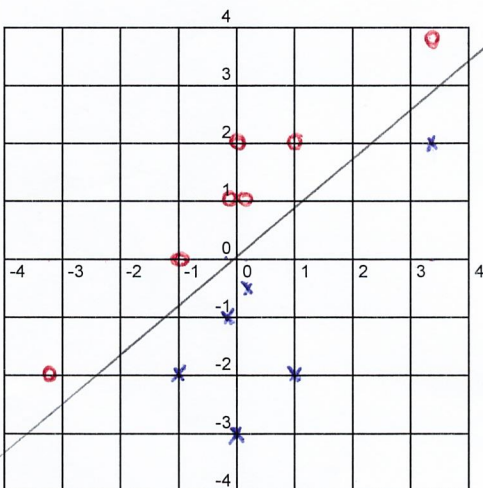
Φ_2

$$\begin{aligned} O &= +1 \\ X &= -1 \end{aligned}$$



Φ_3

$$\begin{aligned} O &= +1 \\ X &= -1 \end{aligned}$$



Bei Φ_3 ist eine Trennung der Daten möglich.