SUPPORT VEKTOR MASCHINEN (SVM) . MASCHINELLES LERNEN

Das Ziel von SVM ist es, die optimale Trennlinie (2D) oder Trennebene (3D) oder Trennhyperebene (mehr als 3D) zu finden, welche die Trainingsdaten (1) optimal (2) trennt.

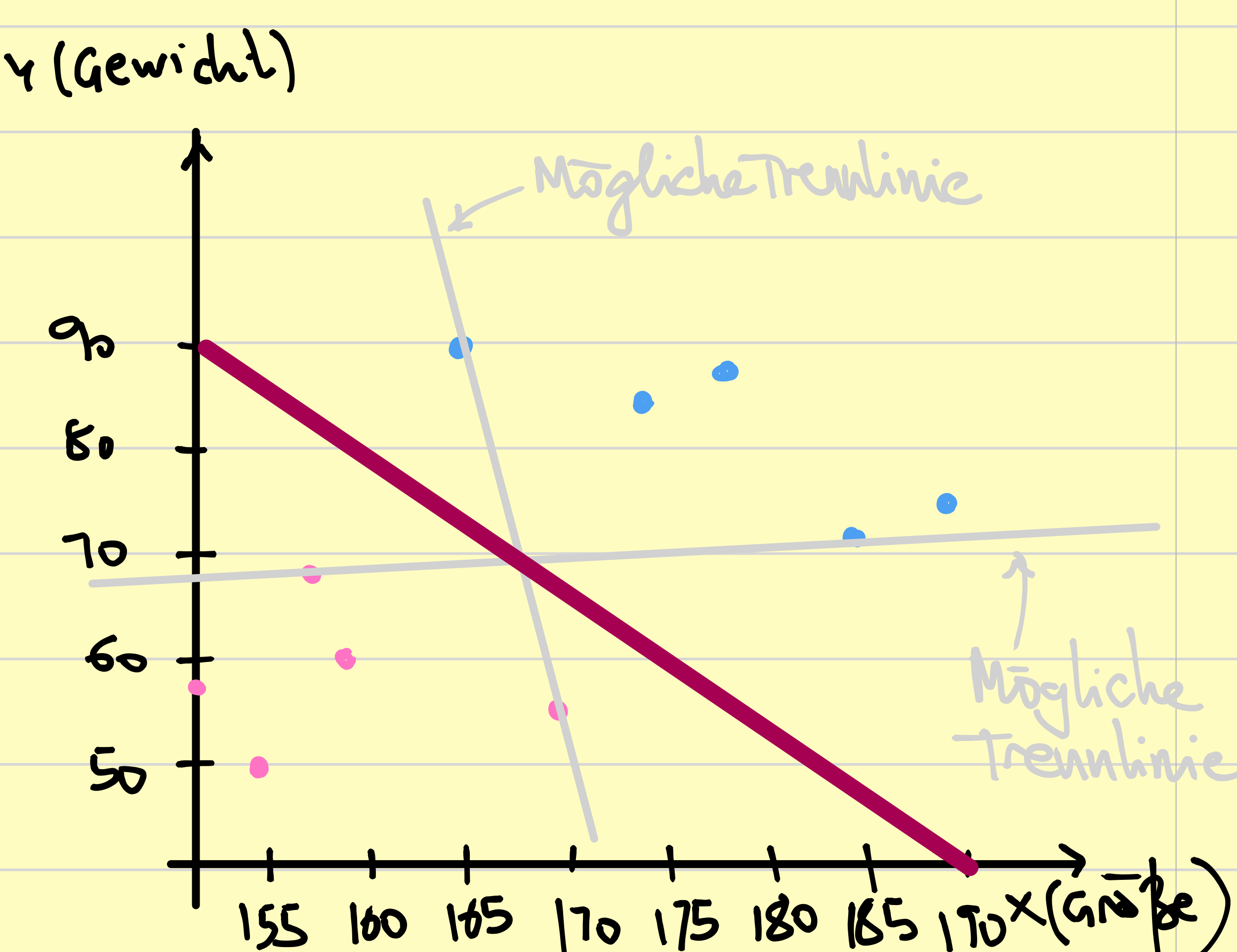
(1) Wir brauchen Trainingsdaten um das SVM Algorythmus anzuwenden. Das heißt, es handelt sich um ein „überwachter" Lernalgorythmus.

(2) Die Trennung der Daten in Klassen/Gruppen (3) ist deshalb optimal, weil der Abstand zu den Daten maximal ist.

(3) SVM ist ein Klassifikationsalgorythmus. Das bedeutet, dass wir die vorhandenen Daten nutzen um hervorzusagen zu welcher Klasse einen neuen Datenpunkt gehört.

BILD·BEISPIEL

X	Y	Klasse
155	50	♀
160	60	♀
158	68	♀
150	58	♀
170	55	♀
165	90	♂
175	85	♂



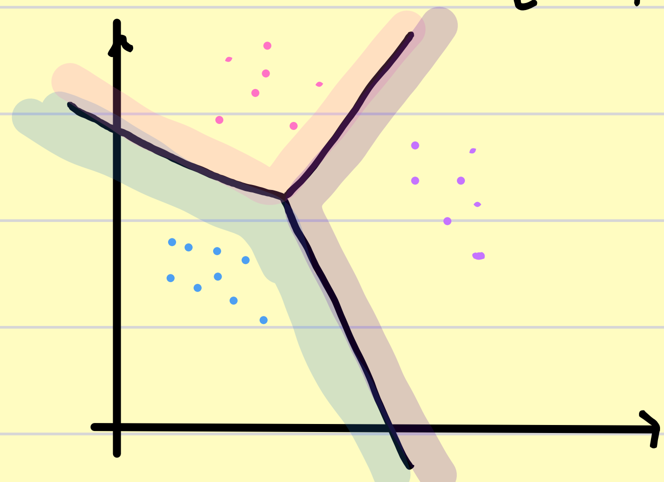
186	88	♂	
190	75	♂	
185	72	♂	

Beide mögliche Trennlinien, trennen die 2D-Fläche in 2 Räume, welche Alle Daten der Klasse beinhalten.

Allerdings der Abstan zw den Trennlinien und den Punkten ist nicht maximal.

Besser ist die Linie durch den Punkten [190,0] [0,90]

Mit 2D und mehrere Klassen:



Beispiel. Gegeben sind folgende Punkte:

ROTE (Klasse A) : [1,1],[2,3],[3,2]

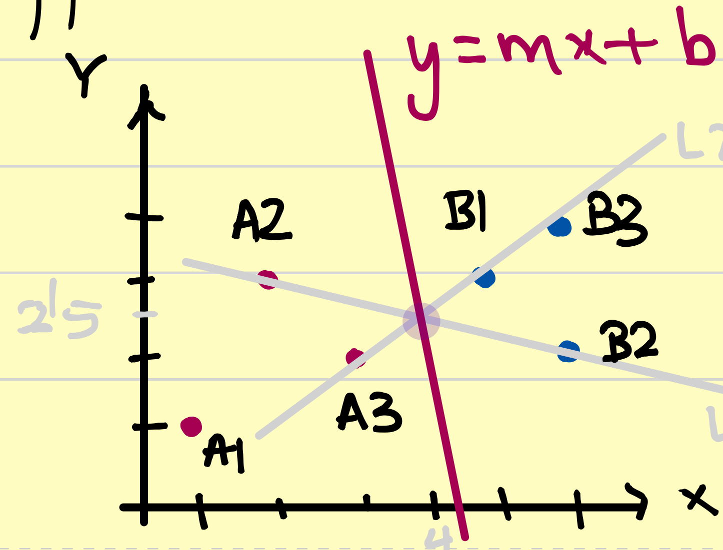
BLAUE (Klasse B) : [5,3],[6,2],[6,4]

Ziel: a) Finden Sie die optimale Trennlinie ($y=mx+b$), welche die beiden Klassen separiert.

b) Berechnen Sie die Abstände (Margin) der nächst. gelegenen Punkten zur Trennlinie.

c) Identifizieren Sie die „Support Vektoren"

Schritt 1. Punkte einzeichnen.

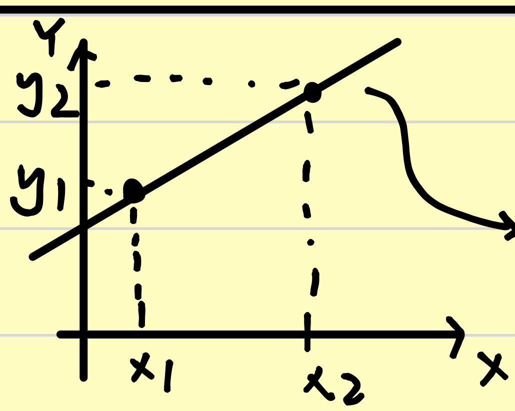


Schritt 2. Form der Trennlinie

$$y = mx + b$$

Schritt 3. Berechnung Punkt (•) der Trennlinie.

L1. $\overline{A2}-\overline{B2}$. $\frac{y-3}{x-2}=\frac{2-3}{6-2} \to \frac{y-3}{x-2} = \frac{-1}{4} \to y = 3-\frac{1}{4}(x-2)$

(•) 

$$\frac{y-y_2}{x-x_2}=\frac{y_2-y_1}{x_2-x_1} \qquad (-)$$

L2. $\overline{A3}-\overline{B1}$ $\frac{y-2}{x-3}=\frac{3-2}{5-3}\to\frac{y-2}{x-3}=\frac{1}{2} \to y=2+\frac{1}{2}(x-3)$

$\to 0 = 1-\frac{1}{4}(x_0-2)-\frac{1}{2}(x_0-3) \to \frac{3}{4} x_0 = 1+\frac{1}{2}+\frac{3}{2}=3$

$\to x_0=4 \to y_0 = 3-\frac{1}{4}(4-2) = 3-\frac{1}{2} = 2'5$

$[x_0,y_0]\to[4, 2'5]$

Schritt 4. Maximierung der Abstände zw. der Linie und die Punkte A3 & B1

Abstand zw. Linie $y=mx+b$ und Punkt. $[x_1,y_1]$

$$d=\frac{|mx_1+y_1+b|}{\sqrt{m^2+1}}$$

Die Linie geht durch den Punkt $[4, 2'5]$

$$y_0 = mx_0 + b \longrightarrow 2'5 = m \cdot 4 + b$$

Abstand Linie und A3: $d_{L \cdot A3} = \dfrac{|m \cdot 3 + 2 + b|}{\sqrt{m^2+1}}$

Abstand Linie und B1: $d_{L \cdot B1} = \dfrac{|m \cdot 5 + 3 + b|}{\sqrt{m^2+1}}$

Beide Abstanden sollten gleich sein, damit der Abstand maximal.

$$3m + 2 + \cancel{b} = 5m + 3 + \cancel{b} \rightarrow 2m = -1 \rightarrow m = \frac{-1}{2}$$

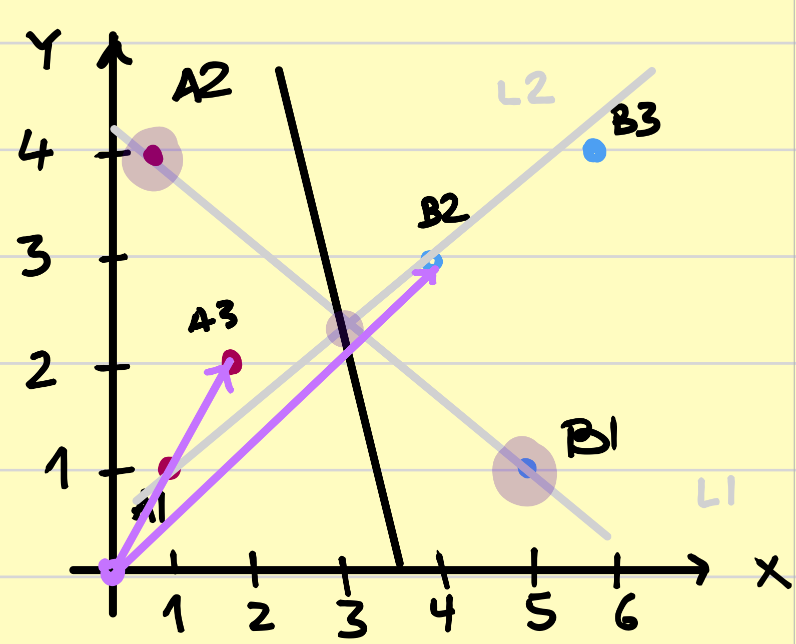
$$b = 0'5$$

$$\boxed{y = -0'5x + 0'5} \quad a)$$

$$d_{L \cdot A3} = \frac{\left|3 \cdot \frac{-1}{2} + 2 + 0'5\right|}{\sqrt{\frac{1}{4}+1}} = \frac{1}{\sqrt{\frac{5}{4}}} = \frac{2}{\sqrt{5}} = \frac{2\sqrt{5}}{5} = d_{L \cdot B1} \quad b)$$

Svektoren A3 $[3,2]$ B1 $[5,3]$ c)

Beispiel.



$$L1: \frac{y-1}{x-5} = \frac{4-1}{1-5} \rightarrow y = 1 - \frac{3}{4}(x-5)$$

$(-)$ $\qquad (-)$

$$L2: \frac{y-3}{x-4} = \frac{1-3}{1-4} \rightarrow y = 3 + \frac{2}{3}(x-4)$$

$$\rightarrow 0 = -2 - \frac{3}{4}(x-5) - \frac{2}{3}(x-4) \rightarrow 2 = \frac{-3}{4}x + \frac{15}{4} - \frac{2}{3}x + \frac{8}{3} \rightarrow$$

$\rightarrow 2 - \frac{15}{4} - \frac{8}{3} = x\left[\frac{-2}{3} - \frac{3}{4}\right] \rightarrow \frac{24-45-32}{12} = x\left[\frac{-8-9}{12}\right] \rightarrow$

$\rightarrow \frac{-53}{12} = x\left[\frac{-17}{12}\right] \rightarrow x_0 = \frac{+53}{+17} = 3'117 \rightarrow y_0 = 1 - \frac{3}{4}\left[3'117-5\right] = 2'41$

$[x_0, y_0] \rightarrow [3'117, 2'41]$

$y = mx + b \rightarrow 2'41 = m \cdot 3'117 + b$

$$\text{Abstand}_{L \cdot A3} = d_{L \cdot A3} = \frac{|m.2+2+b|}{\sqrt{m^2+1}}$$

$$\text{Abstand}\ L \cdot B2 = d_{L \cdot B2} = \frac{|m.4+3+b|}{\sqrt{m^2+1}}$$

$d_{L \cdot A3} = d_{L \cdot B2} \rightarrow 2m+2+\cancel{b} = 4m+3+\cancel{b} \rightarrow 2m = -1 \rightarrow m = -0'5$

$\rightarrow 2'41 = -0'5 \cdot 3'117 + b \rightarrow b = 2'41 + 0'5 \cdot 3'117 = 3'9585$

$$\boxed{y = -0'5\,x + 3'9585}$$

$$\text{Abstand}\ L \cdot A3: \quad d = \frac{|2.(-0'5+2+3'9585)|}{\sqrt{0'5^2+1}} =$$

Svcktoren: $A3[2,2]$ $B2[4,2]$