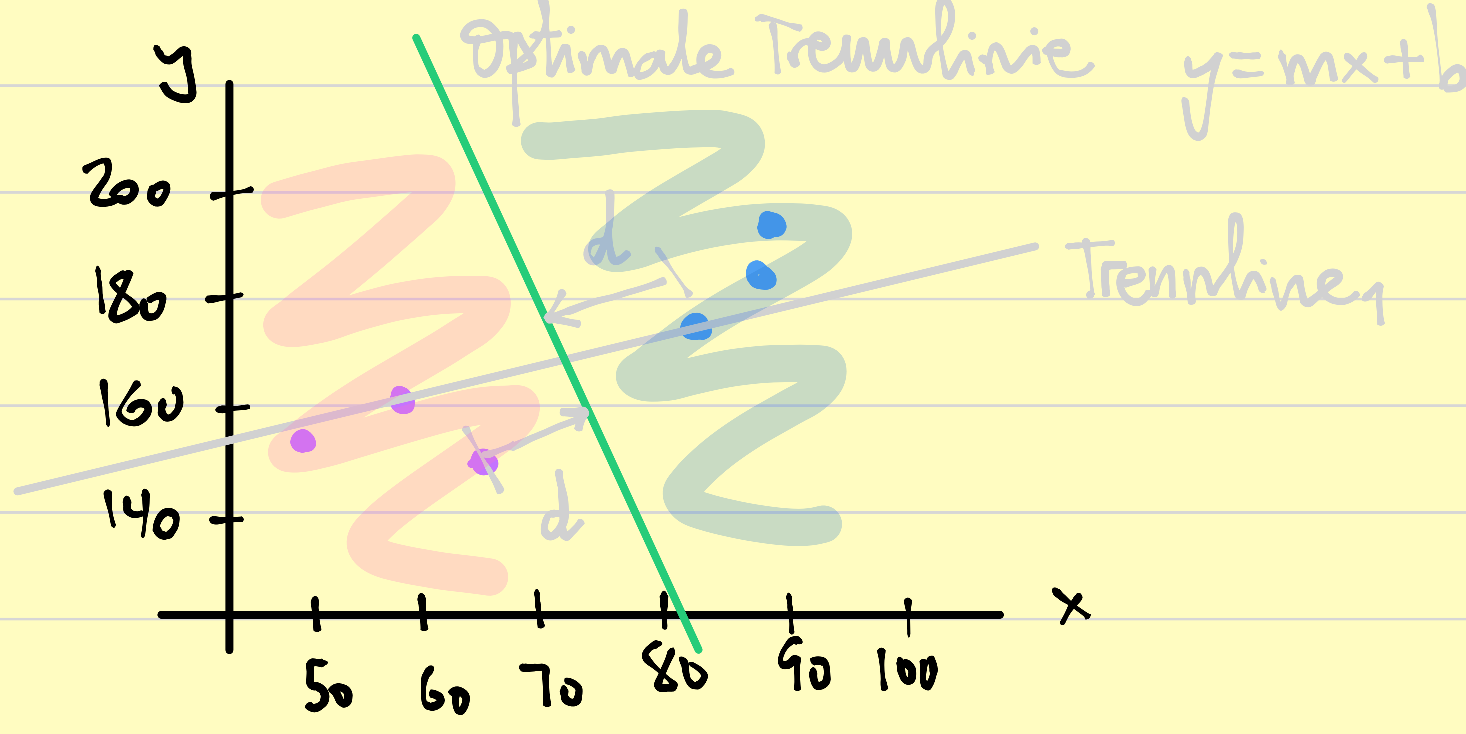
Support Vektor Maschines (SVM)

Das Ziel von SVM ist die optimale Trennlinie zu finden, welche Trainingsdaten teilt.

- Wir brauchen dafür gelabelte Daten.
- Die Trennung der Klassen erfolgt über Abstände. ? Normierung ?
- Die Klassifizierung erfolgt anhand bereits bekannte Daten.

Beispiel.

x (Gewicht)	y (Größe)	Klasse
50	155	♀
60	160	♀
68	150	♀
85	175	♂
90	182	♂
92	192	♂



Beispiel. Gegeben sind folgende Punkte:

ROTE (A): $[1,1]_{A1}$, $[2,3]_{A2}$, $[3,2]_{A3}$ BLAUE (B): $[5,3]_{B1}$, $[6,2]_{B2}$, $[6,4]_{B3}$

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

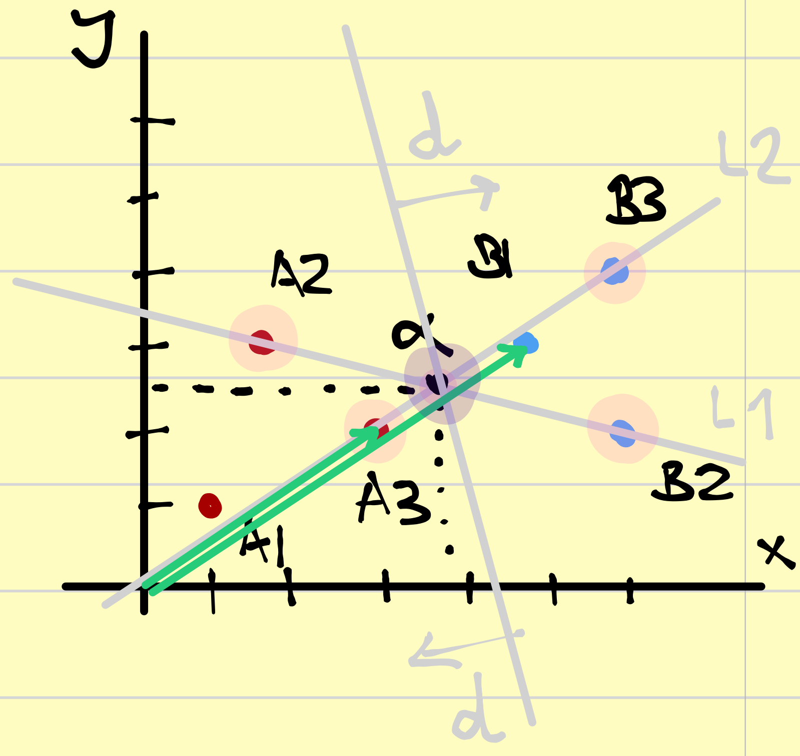
b) Berechnen Sie die Abstände (d) der nächstgelegenen Punkten zur Trennlinie.

c) Identifizieren Sie die Support Vektoren.

Schritt 1. Punkte Visualisieren

Schritt 2. Form der Trennlinie $y=mx+b$

Schritt 3. Extrem Trennlinien ermitteln.



$$\boxed{\frac{y-y_2}{x-x_2}=\frac{y_1-y_2}{x_1-x_2}}$$

EINE LINIE WELCHE DURCH $[x_1 y_1]$ & $[x_2 y_2]$ GEHT.

$$L1: \frac{y-2}{x-6}=\frac{3-2}{2-6} \longrightarrow$$

$$y-2=\frac{-1}{4}(x-6)$$

$$L2: \frac{y-4}{x-6}=\frac{2-4}{3-6}$$

$$y-4=\frac{2}{3}(x-6)$$

$$(y_\alpha-2)-(y_\alpha-4)=\frac{-1}{4}(x_\alpha-6)-\frac{2}{3}(x_\alpha-6) \rightarrow$$

$$\rightarrow 2=\frac{-11}{12}(x_\alpha-6) \rightarrow x_\alpha=3'818 \longrightarrow y_\alpha=2'545$$

Die optimale Trennlinie geht durch α $[3'818, 2'545]$

$$y_\alpha = m x_\alpha + b \quad (1)$$

Schritt 4. d ermitteln.

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

ABSTAND zu linie $y = mx+b$
UND PUNKT $[x_1, y_1]$

d : linie A3 : $$d = \frac{|m.3 + 2 + b|}{\sqrt{m^2+1}} \quad (2)$$

d : Linie B1 : $$d = \frac{|m.5 + 3 + b|}{\sqrt{m^2+1}} \quad (3)$$

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

$$y_\alpha = m . x_\alpha + b \longrightarrow 2'545 = \frac{-1}{2} . 3'818 + b \rightarrow b = 0'636$$

$\alpha = [3'818, 2'545]$

$$\boxed{y = \frac{-1}{2} x + 0'636}$$

$$d = \frac{\left| \frac{-1}{2} . 5 + 3 + 0'636 \right|}{\sqrt{\left(\frac{-1}{2}\right)^2 + 1}} = 1'016 \rightarrow \boxed{d = 1'016}$$

Support Vektoren sind die Punkte A3 & B1.

Übung: bitte ermitteln Sie die optimale Trennlinie und die Support Vektoren um die o.g. Daten Gewicht · Größe optimal zu trennen.