

Classification II

Daten: $(x_i, y_i) \in \mathbb{R}^p \times \{\pm 1\}$
 Merkmale $\nearrow$ Klasse
 $i = 1, \dots, n$



$x_i \in C_1$

"kreditwürdig" $\Rightarrow y_i = +1$

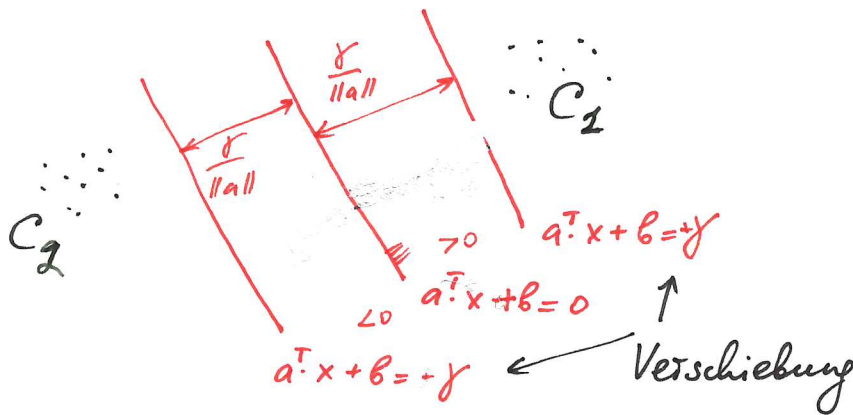
$x_i \in C_2$

"nicht kreditwürdig" $\Rightarrow y_i = -1$

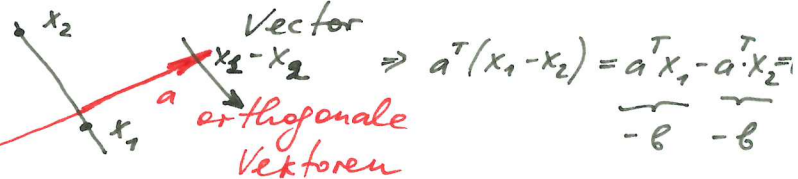
Trennende Hyperebene: $a^T x + b = 0$
 (Angenommen, existent)

Support Vector Machine (SVM):

Wähle $a \in \mathbb{R}^p$ und $b \in \mathbb{R}$ so, daß der Abstand ^{minimales} zu beider Klassen C_1 und C_2 maximal wird. ^{Min Max -Ansatz}



• Hyperebene: $a^T x + b = 0$



• Abstand zwischen $a^T x + b_1 = 0$ & $a^T x + b_2 = 0$

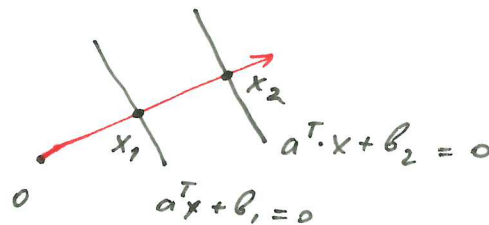
$$x_1 = d_1 \cdot a, \quad x_2 = d_2 \cdot a$$

$$\Rightarrow a^T (d_1 \cdot a) + b_1 = 0 \Rightarrow d_1 = -\frac{b_1}{\|a\|^2}$$

$$a^T (d_2 \cdot a) + b_2 = 0 \Rightarrow d_2 = -\frac{b_2}{\|a\|^2}$$

Abstand

$$\|x_1 - x_2\| = \left\| -\frac{b_1}{\|a\|^2} a + \frac{b_2}{\|a\|^2} a \right\| = \frac{|b_1 - b_2|}{\|a\|}$$



$$\max_{a, b, \gamma \geq 0} \frac{\gamma}{\|a\|}$$

s.d. $a^T x_i + b \geq \gamma, i \in C_1 \Leftrightarrow y_i (a^T x_i + b) \geq \gamma, i = 1, \dots, n$
 $a^T x_i + b \leq -\gamma, i \in C_2$

$$\min_{a, b, \gamma \geq 0} \frac{\|a\|}{\gamma}$$

s.d. $y_i \left(\frac{a^T x_i + b}{\gamma} \right) \geq \frac{\gamma}{\gamma}, i = 1, \dots, n$

neue skalierte Variablen

$$\min_{a \in \mathbb{R}^p, b \in \mathbb{R}} \frac{1}{2} \|a\|^2 \text{ s.d. } y_i (a^T x_i + b) \geq 1, i=1, \dots, n \quad (2)$$

↑ Quadratisches Optimierungsproblem mit n linearen Nebenbedingungen

→ optimaler Abstandsklassifikator a^*, b^*

x gehört zu C_1 , falls $a^{*T} x + b^* > 0$
 x gehört zu C_2 , falls $a^{*T} x + b^* < 0$
 "Newcomer"

Lagrange-Multiplikator $\rightarrow +\infty$, da d kann groß gewählt werden
 $\rightarrow$ zu min a, b

Duales Problem:

vertauschen möglich

"starke Dualität, da lin. NB"

$$\min_{a, b} \max_{d \geq 0} \frac{1}{2} \|a\|^2 - \sum_{i=1}^n d_i \cdot [y_i (a^T x_i + b) - 1]$$

$$\min_{a, b} \frac{1}{2} \|a\|^2 - \sum_{i=1}^n d_i \cdot y_i (a^T x_i + b) + \sum_{i=1}^n d_i$$

Optimalität $\nabla_a = 0, \nabla_b = 0$: $a = \sum_{i=1}^n d_i \cdot y_i \cdot x_i$; $\sum_{i=1}^n d_i \cdot y_i = 0$.

$$\left(= \frac{1}{2} \|a\|^2 - a^T \sum_{i=1}^n d_i \cdot y_i \cdot x_i - \sum_{i=1}^n d_i \cdot y_i \cdot b = -\frac{1}{2} \|a\|^2 \right)$$

$$\max_{d \geq 0} \sum_{i=1}^n d_i - \frac{1}{2} \sum_{i,j=1}^n d_i \cdot d_j \cdot y_i \cdot y_j \cdot x_i^T \cdot x_j \text{ s.d. } \sum_{i=1}^n d_i \cdot y_i = 0$$

($d \in \mathbb{R}^n$)
 n Variablen

Quadratisches Optimierungsproblem mit einer Nebenbedingung

→ λ^* optimaler Lagrange-Multiplikator

Recovery:

$$a^* = \sum_{i \in S} d_i^* \cdot y_i \cdot x_i \quad \& \quad d_i \neq 0 \Rightarrow y_i (a^{*T} x_i + b) - 1 = 0 \quad [: y_i]$$

nicht verschwindet

Lagrange-Multiplikatoren Ungleichung ist aktiv wegen Komplementarität

$$d_i > 0 \Rightarrow y_i (a^{*T} x_i + b) = 1$$

↑ Stützpunkte mit minimalem Abstand zu der Hyperebene $a^{*T} x + b^* = 0$

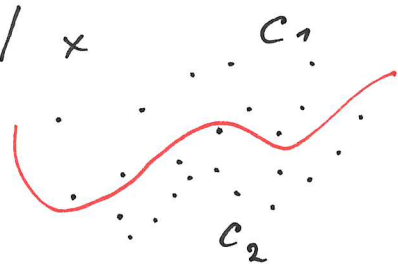
$$d(x) := a^{*T} x + b^* = \sum_{i \in S} d_i^* \cdot y_i \cdot x_i^T \cdot x + b^* \begin{cases} > 0 \Rightarrow y = +1 (C_1) \\ < 0 \Rightarrow y = -1 (C_2) \end{cases}$$

nur Skalarprodukte $x_i^T \cdot x$ spielen eine Rolle, $i \in S$

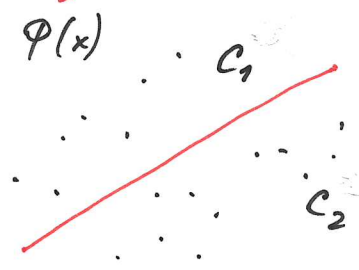
Trennbarkeit: Kernel

(3)

Merkmal x



Feature $\Phi(x)$



x gehört zu $C_j \Leftrightarrow \Phi(x)$ gehört zu C_j
 $j=1,2$

→ Trennung von $\Phi(x_i), i=1, \dots, n$ mit Hilfe von SVM ↓ dualer Ansatz

$$\max_{d \geq 0} \sum_{i=1}^n d_i - \frac{1}{2} \sum_{i,j=1}^n d_i d_j y_i y_j \underbrace{\Phi^T(x_i) \cdot \Phi(x_j)}_{= K(x_i, x_j) - \text{Kernel}} \text{ s.d. } \sum_{i=1}^n d_i y_i = 0$$

z.B. • $K(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}}$ Gauß-Kernel

• $K(x, z) = (x^T \cdot z)^2$ Euklidischer Kernel

Intuition: $\Phi(x) \approx \Phi(z) \Rightarrow \Phi^T(x) \cdot \Phi(z)$ groß
 $\Phi(x) \perp \Phi(z) \Rightarrow \Phi^T(x) \cdot \Phi(z) = 0$

Ähnlichkeit von x und z

Frage: Wann ist $K(x, z) = \Phi^T(x) \cdot \Phi(z)$ für $\{x_1, \dots, x_n\}$ darstellbar?

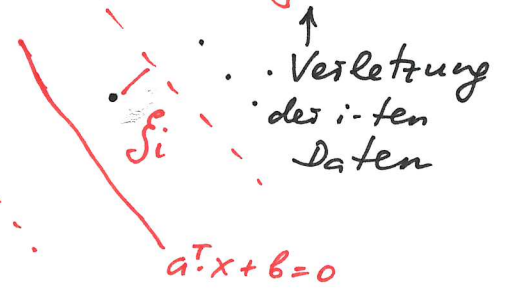
(Mercier: $(K(x_i, x_j))_{i,j}$ - symmetrisch & positive semi-definite.)

Trennbarkeit: Regularisierung

$$\min_{a, b, f \geq 0} \frac{1}{2} \|a\|^2 + c \cdot \sum_{i=1}^n f_i \text{ s.d. } y_i (a^T x + b) \geq 1 - f_i, i=1, \dots, n$$

Bestrafung des Fehlers

↓ Duales Problem: (ii)



$$\max_{c \geq d \geq 0} \sum_{i=1}^n d_i - \frac{1}{2} \sum_{i,j=1}^n d_i d_j y_i y_j x_i^T x_j \text{ s.d. } \sum_{i=1}^n d_i y_i = 0$$

Lagrange-Multiplikatoren sind beschränkt

