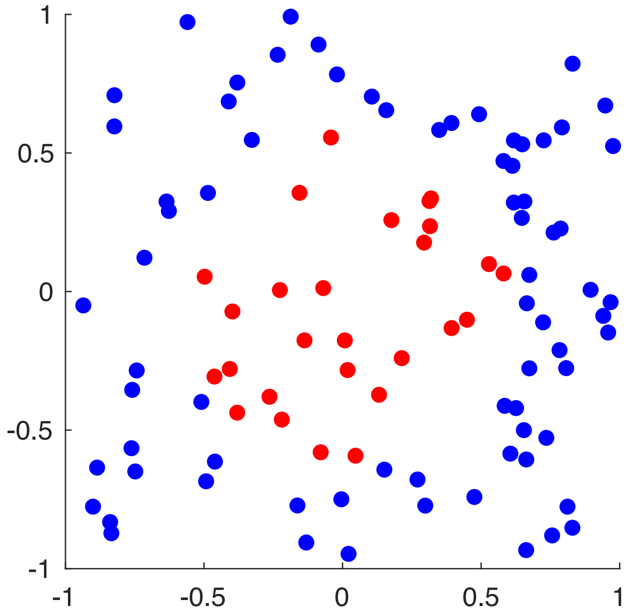
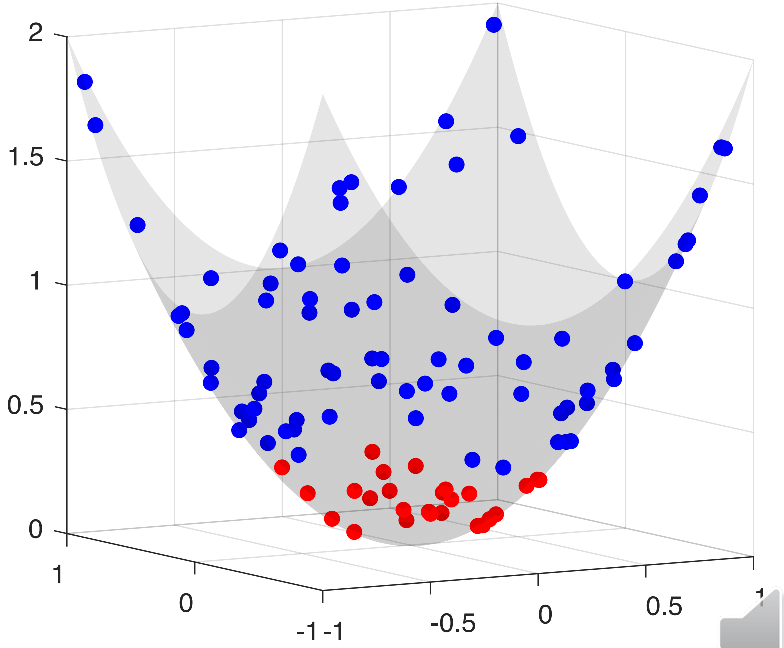
5c. Lineare Nebenbedingungen

Lagrange-Dualität



Optimierung SoSe 2020
Dr. Alexey Agaltsov



Plan

- Lagrange-Dualität
- Dualität für restringierte Optimierungsprobleme
- Beispiel: Support Vector Machines

Primales und duales Problem

- Sei $\mathcal{L}: A \times B \to \mathbb{R}$ gegeben
- Als primales Problem bezeichnen wir:

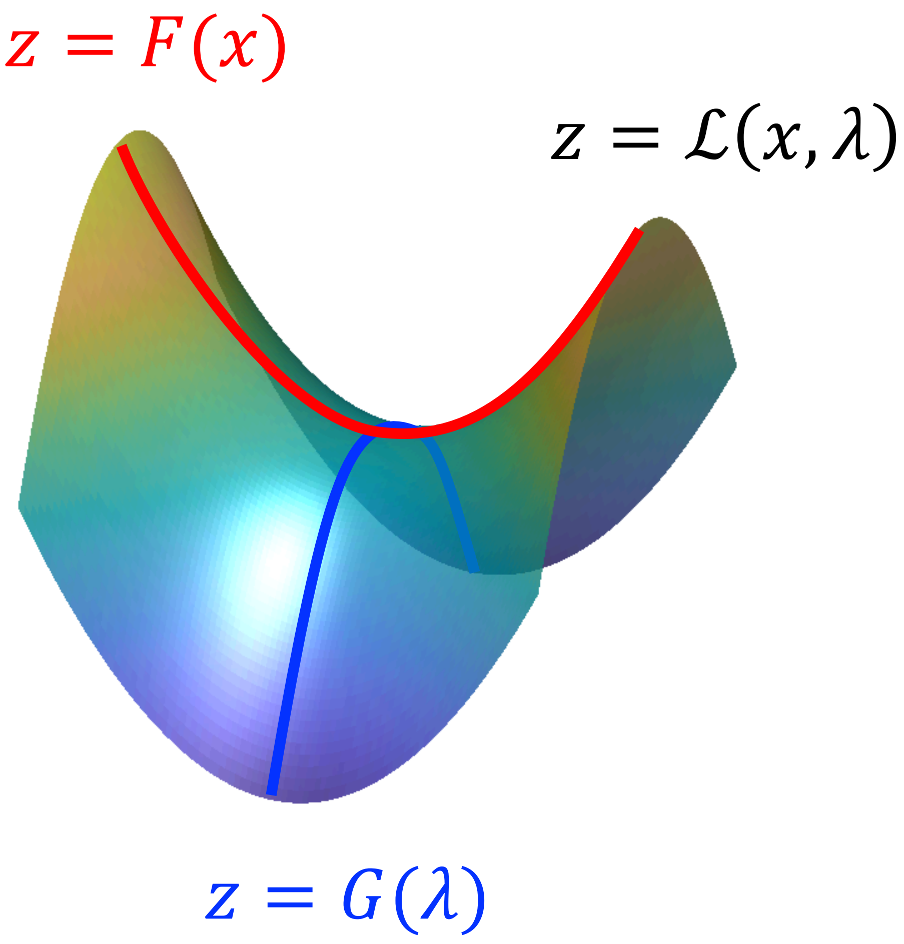
$$\text{Minimiere } F(x), \quad x \in A$$
$$(\text{u.d.N. } F(x) < \infty)$$

Primale Funktion $\quad F(x) = \sup_{\lambda \in B} \mathcal{L}(x, \lambda)$

- Als duales Problem bezeichnen wir:

$$\text{Maximiere } G(\lambda), \quad \lambda \in B$$
$$(\text{u.d.N. } G(\lambda) > -\infty)$$

Duale Funktion $\quad G(\lambda) = \inf_{x \in A} \mathcal{L}(x, \lambda)$



Satz 5.13. Schwache Dualität

Seien F, G die primale bzw. die duale Funktionen. Dann gilt:

$$G(\lambda) \leq F(x) \quad \forall x \in A, \lambda \in B$$

Beweis:

$$\underbrace{\inf_{y \in A} \mathcal{L}(y, \lambda)}_{G(\lambda)} \leq \mathcal{L}(x, \lambda) \leq \underbrace{\sup_{u \in B} \mathcal{L}(x, u)}_{F(x)} \quad \forall x \in A, \forall \lambda \in B$$

$\square$

ε-Suboptimalität

$$F_* = \inf_{x \in A} F(x)$$

- Ein Punkt $x \in A$ heißt ε-suboptimal falls:

$$F(x) - F_* < \varepsilon$$

- Jeder Punkt $\lambda \in B$ entspricht einer Abschätzung:

$$F_* \geq G(\lambda)$$

- Seien $x \in A, \lambda \in B$ so, dass $F(x) - G(\lambda) < \varepsilon.$ Dann ist x ε-suboptimal

Starke Dualität

$$F_* = \inf_{x \in A} F(x), \quad G_* = \sup_{\lambda \in B} G(\lambda)$$

- Als Dualitätslücke bezeichnen wir:

$$F_* - G_* \geq 0$$

- Die starke Dualität ist erfüllt, falls:

$$F_* = G_*$$

Beispiel: Positive Dualitätslücke

- Sei $\mathcal{L}(x, \lambda)$ in der folgenden Tabelle aufgeführt:

$x \setminus \lambda$	1	2
1	1	0
2	0	1

$$F(x) = \sup_{\lambda \in \{1,2\}} \mathcal{L}(x, \lambda) \equiv 1$$

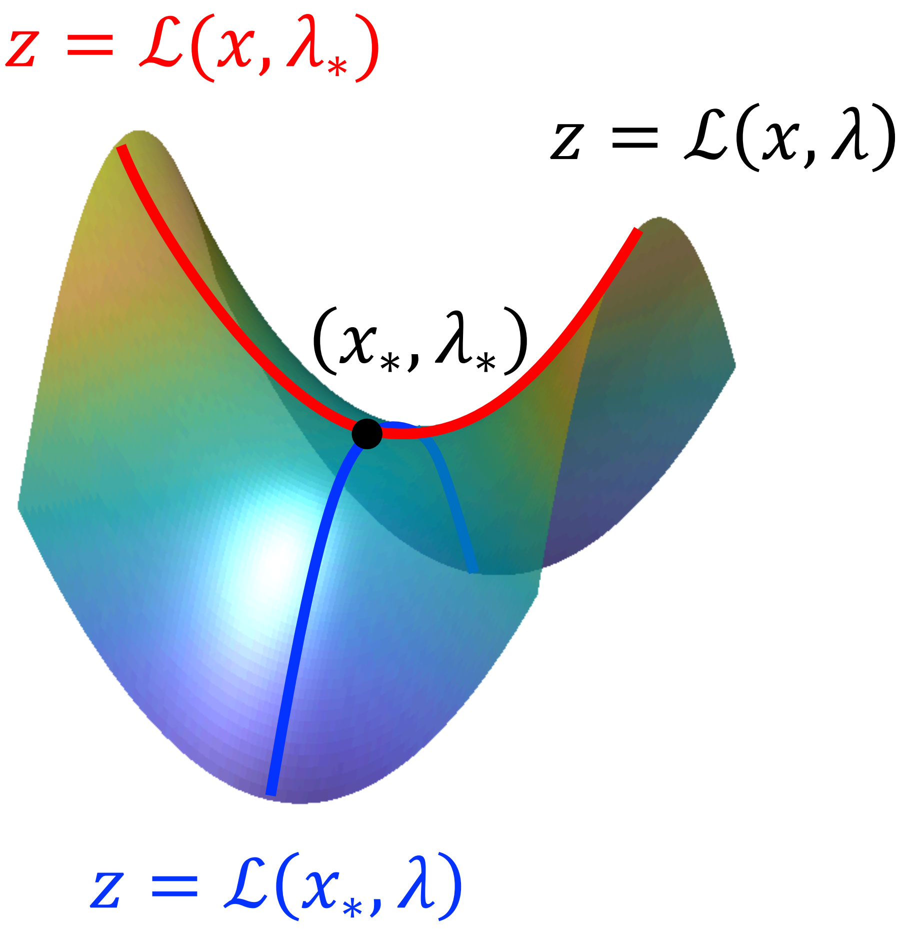
$$G(\lambda) = \inf_{x \in \{1,2\}} \mathcal{L}(x, \lambda) \equiv 0$$

$$F_* - G_* = 1$$

Sattelpunkte

$(x_*, \lambda_*) \in A \times B$ heißt Sattelpunkt von $\mathcal{L}$ falls:

$$\mathcal{L}(x, \lambda_*) \leq \mathcal{L}(x_*, \lambda_*) \leq \mathcal{L}(x, \lambda_*) \quad \forall x \in A, \lambda \in B$$



Satz 5.14. Starke Dualität

Die folgenden Aussagen sind äquivalent:

- (x_*, λ_*) ist ein Sattelpunkt von $\mathcal{L}$
- $x_* \in \mathrm{Argmin}_x(F),\ \lambda_* \in \mathrm{Argmax}_\lambda(G)$ und $G(\lambda_*) = F(x_*)$

Beweis

$$\mathcal{L}(x_*, \lambda) \le \mathcal{L}(x_*, \lambda_*) \le \mathcal{L}(x, \lambda_*) \qquad \forall x \in A, \lambda \in B$$

$$\underset{\substack{\| \\ F(x_*)}}{\sup_{\lambda \in B} \mathcal{L}(x_*, \lambda)} = \mathcal{L}(x_*, \lambda_*) = \underset{\substack{\| \\ G(\lambda_*)}}{\inf_{x \in A} \mathcal{L}(x, \lambda_*)}$$

schwache Dualität

$$x_* \in \operatorname{Argmin}_x(F),\ \lambda_* \in \operatorname{Argmax}_\lambda(G),\ F(x_*) = G(\lambda_*)$$

$\square$

Anwendung der starken Dualität

- Angenommen, das primale und das duale Probleme besitzen optimale Lösungen x_*, λ_* und die starke Dualität ist erfüllt

- Man kann x_* durch die Lösung eines Problems erhalten, die in der Regel einfacher als das primale Problem ist:

$$x_* = argmin_x \mathcal{L}(x, \lambda_*)$$

- Oft ist das duale Problem einfacher zu lösen, als das primale Problem

Plan

- Sattelpunkte und Dualität
- Dualität für restringierte Optimierungsprobleme
- Beispiel: Support Vector Machines

Restringiertes Optimierungsproblem

beliebige Grundmenge

Minimiere $f(x)$ über $x \in A$

u.d.N. $g(x) \leq 0$

$h(x) = 0$

$f: A \to \mathbb{R}$

$g: A \to \mathbb{R}^{m_U}$

$h: A \to \mathbb{R}^{m_G}$

- Dieses Problem lässt sich als primales Problem im Rahmen der Lagrange-Dualität formulieren
- Als Lagrange-Funktion bezeichnen wir:

≤ 0 *für zulässiges* x

$$\mathcal{L}(x, \mu, \lambda) := f(x) + \overbrace{\underset{\geq 0}{\mu}^T g(x)} + \lambda^T h(x)$$

$$x \in A,\ (\mu, \lambda) \in B = \mathbb{R}^{m_U}_{\geq 0} \times \mathbb{R}^{m_G}$$

Das primale Problem

$$\mathcal{L}(x,\mu,\lambda) \coloneqq f(x) + \mu^T g(x) + \lambda^T h(x)$$

$$F(x) = \sup_{\mu \geq 0, \lambda} \mathcal{L}(x,\mu,\lambda) = \begin{cases} f(x) & h(x) = 0, g(x) \leq 0 \\ \infty & h(x) \neq 0 \text{ oder } g(x) \nleq 0 \end{cases}$$

Minimiere $f(x)$ über $x \in A$

u.d.N. $g(x) \leq 0$

$h(x) = 0$

$\Leftrightarrow$ Minimiere $F(x)$ über $x \in A$

u.d.N. $F(x) < \infty$

Das duale Problem

$$\mathcal{L}(x,\mu,\lambda) = f(x) + \mu^T g(x) + \lambda^T h(x)$$

$$G(\mu,\lambda) = \inf_{x\in A} \underbrace{\mathcal{L}(x,\mu,\lambda)}_{\text{linear in } \mu,\lambda} \xrightarrow{\text{Satz 2.12}} \text{konkav}$$

Maximiere $G(\mu,\lambda)$ über $(\mu,\lambda) \in \mathbb{R}^{m_U} \times \mathbb{R}^{m_G}$

u.d.N. $\mu \geq 0$

Konkaves Maximierungsproblem

sogar wenn das primale Problem nicht konvex ist

Satz 5.15. Starke Dualität für lineare NB

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $A_U x \geq b_U$

$A_G x = b_G$

$f \in C^1(\mathbb{R}^n)$ konvex

$A_U \in \mathbb{R}^{m_U \times n}, b_U \in \mathbb{R}^{m_U}$

$A_G \in \mathbb{R}^{m_G \times n}, b_G \in \mathbb{R}^{m_G}$

Sei x_* eine optimale Lösung und μ_*, λ_* die Lagrange-Multiplikatoren

Dann ist (μ_*, λ_*) eine optimale Lösung des dualen Problems

Außerdem stimmen die optimalen Funktionswerte des primalen und des dualen Problems überein

Also gilt die starke Dualität

Beweis

Satz 5.12 (KKT)

$\mu_* \in \mathbb{R}_+^{m_U}, \lambda_* \in \mathbb{R}^{m_G}$ sind so, dass:

$$\nabla_x \mathcal{L}(x_*, \mu_*, \lambda_*) = 0$$

$\mathcal{L}(\cdot, \mu_*, \lambda_*)$ ist konvex

Satz 2.20

$$\mathcal{L}(x_*, \mu_*, \lambda_*) \leq \mathcal{L}(x, \mu_*, \lambda_*) \quad \forall x \in \mathbb{R}^n$$

$\forall \mu \geq 0, \lambda$: $\quad \mathcal{L}(x_*, \mu, \lambda) = f(x_*) + \underbrace{\mu^T}_{\geq 0}\overbrace{(b_U - A_U x_*)}^{\leq 0} + \lambda^T \overbrace{(b_G - A_G x_*)}^{= 0}$

$$\leq f(x_*) = \mathcal{L}(x_*, \mu_*, \lambda_*)$$

$\mu_{*,i}(b_{U,i} - A_{U,i} x_*) = 0$ *(Satz 5.12)*

$\Rightarrow$ (x_*, μ_*, λ_*) ist ein Sattelpunkt von $\mathcal{L}$

Satz 5.14

(μ_*, λ_*) ist ein optimaler Dualpunkt und die starke Dualität ist erfüllt

$\square$

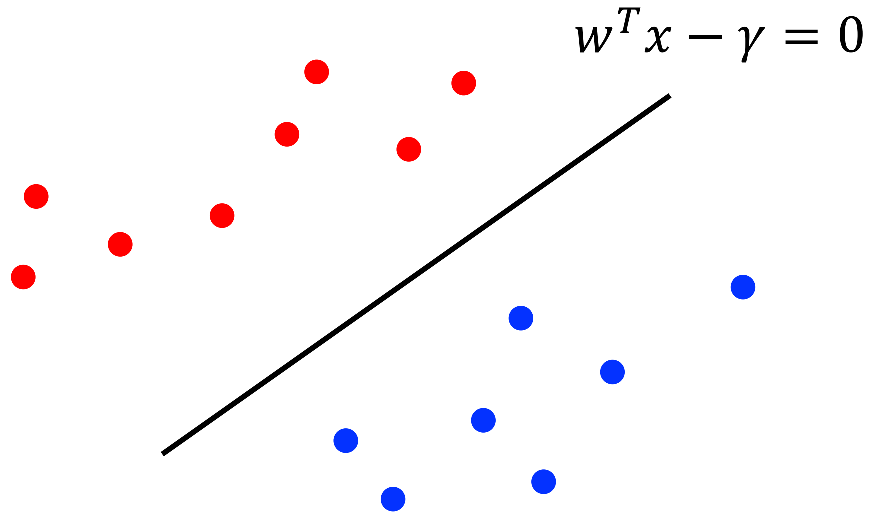
Plan

- Lagrange-Dualität
- Dualität für restringierte Optimierungsprobleme
- Beispiel: Support Vector Machines

Binäre Klassifizierung

- Seien $x_1, \ldots, x_m \in \mathbb{R}^n$ Merkmalsvektoren
- Seien $t_1, \ldots, t_m \in \{1, -1\}$ die zugehörigen Klassen
- Im einfachsten Fall sind die Klassen linear separierbar

$$w^T x_i - \gamma > 0 \;\Leftrightarrow\; t_i = 1$$
$$w^T x_i - \gamma < 0 \;\Leftrightarrow\; t_i = -1$$
$$w \in \mathbb{R}^n \setminus \{0\}, \gamma \in \mathbb{R}$$



Binäre Klassifizierung

$$w^T x_i - \gamma > 0 \;\Leftrightarrow\; t_i = 1$$
$$w^T x_i - \gamma < 0 \;\Leftrightarrow\; t_i = -1$$

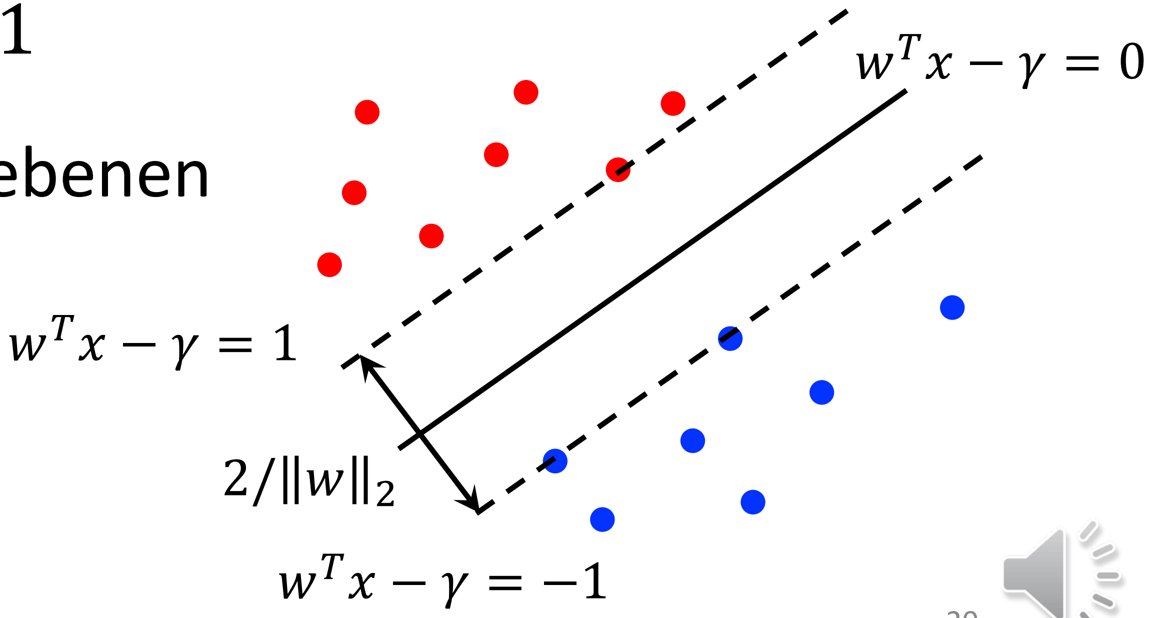
Eventuell $(w, \gamma) := (w/a, \gamma/a)$

$$a = \min_i |w^T x_i - \gamma|$$

$$w^T x_i - \gamma \geq 1 \;\Leftrightarrow\; t_i = 1$$
$$w^T x_i - \gamma \leq -1 \;\Leftrightarrow\; t_i = -1$$

- Die Entfernung zwischen den Hyperebenen $w^T x - \gamma = \pm 1$ ist:

$$2/\|w\|_2$$

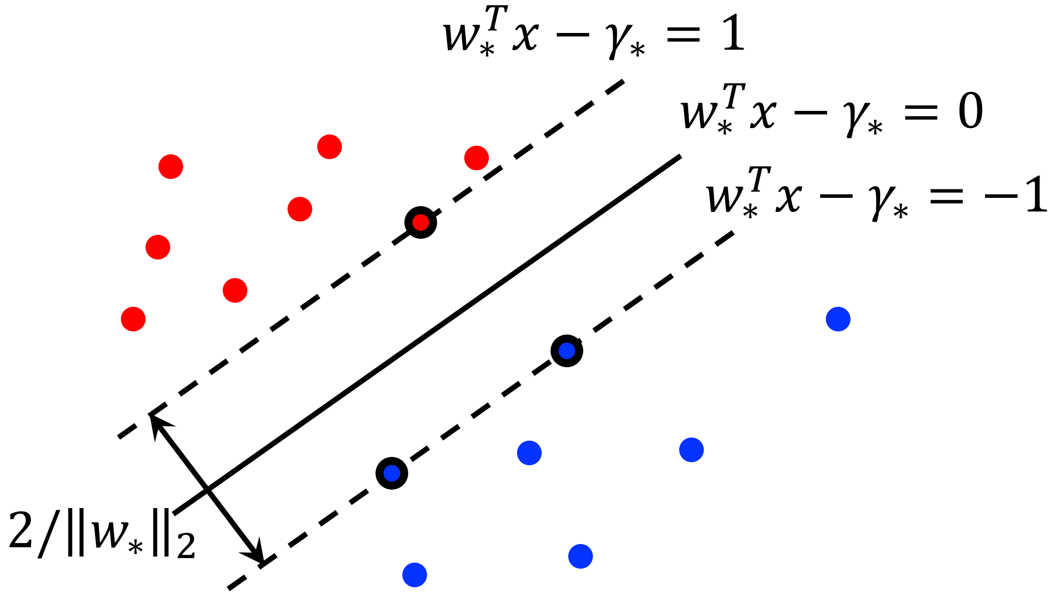


Support Vector Machine

- SVM: Wähle w, γ so, dass der Abstand $2/\|w\|_2$ maximal ist

  Minimiere $\frac{1}{2}\|w\|_2^2$ über w, γ

  u.d.N. $t_i(w^T x_i - \gamma) \geq 1,\ i = 1, \ldots, m$



- Sei w_*, γ_* eine optimale Lösung
- x_k heißt Stützvektor, falls auf einer Grenzhyperebene liegt:

$$w_*^T x_k - \gamma_* = \pm 1$$

Aufgabe 5.16. Support Vector Machine

$$\text{Minimiere } \frac{1}{2}\|w\|_2^2 \text{ über } w \in \mathbb{R}^n, \gamma \in \mathbb{R}$$

$$\text{u.d.N. } t_i\left(x_i^T w - \gamma\right) \geq 1, \quad i = 1, \ldots, m$$

- Der Zulässigkeitsbereich ist nichtleer $\Leftrightarrow$ die Klassen sind linear separierbar

- Die Klassen sind nichtleer und linear separierbar $\Rightarrow$ das primale und das duale Probleme haben optimale Lösungen und die starke Dualität ist erfüllt

Tipp: Verweisen sie auf Satz 5.15

Die duale Funktion

- Lagrange-Funktion:

$$\mathcal{L}(w,\gamma;\mu) = \frac{1}{2}w^T w + \sum_{i=1}^m \mu_i(1 - t_i(w^T x_i - \gamma))$$

- Die duale Funktion:

$$G(\mu) = \inf_{w,\gamma} \mathcal{L}(w,\gamma;\mu)$$

- $\mathcal{L}(\cdot,\cdot;\mu)$ ist quadratisch für ein festes μ. Optimalitätsbedingungen:

$$\begin{aligned} \nabla_w \mathcal{L}(w,\gamma;\mu) &= w - \sum_{i=1}^m \mu_i t_i x_i = 0 \\ \nabla_\gamma \mathcal{L}(w,\gamma;\mu) &= \sum_{i=1}^m \mu_i t_i = 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} w_* &= \sum_{i=1}^m \mu_i t_i x_i \\ \sum_{i=1}^m \mu_i t_i &= 0 \end{aligned}$$

Aufgabe 2.21

$$G(\mu) = \begin{cases} \mathcal{L}(w_*,\gamma,\mu) = \sum_{i=1}^m \mu_i - \frac{1}{2}\sum_{i,j=1}^m \mu_i \mu_j t_i t_j x_i^T x_j, & \sum_{i=1}^m \mu_i t_i = 0,\ \forall\gamma \\ -\infty, & \text{sonst} \end{cases}$$

Das duale Programm

$$\begin{aligned} \text{Maximiere} \quad & \sum_{i=1}^{m} \mu_i - \frac{1}{2}\sum_{i,j=1}^{m} \mu_i \mu_j t_i t_j x_i^T x_j \quad \text{über } \mu_1, \dots, \mu_m \in \mathbb{R} \\ \text{u.d.N.} \quad & \sum_{i=1}^{m} \mu_i t_i = 0 \\ & \mu_1, \dots, \mu_m \geq 0 \end{aligned}$$

- Sei w_*, γ_* eine optimale Lösung des primalen Problems und sei μ_* der zugehörige Vektor der Lagrange-Multiplikatoren

μ_* *ist eine optimale Lösung des dualen Problems nach Satz 5.15*

$(w_*, \gamma_*) = argmin_{w,\gamma}\, \mathcal{L}(w, \gamma, \mu_*)$ *nach Satz 5.14* $\Rightarrow w_* = \sum_{i=1}^{m} \mu_{*,i} t_i x_i$

- Komplementaritätsbedingung (Satz 5.12):

$\mu_{*,i}\big(1 - t_i(w_*^T x_i - \gamma_*)\big) = 0 \;\; \Rightarrow \;\; \mu_{*,i} > 0$ nur wenn x_i ein Stützvektor ist

Sei x_i ein Stützvektor $\Rightarrow \gamma_* = w_*^T x_i - t_i$

Klassifizierungsregel

- Sei $x \in \mathbb{R}^n$. Dann gehört x zur Klasse

$$t(x) = \operatorname{sgn}(w_*^T x - \gamma_*)$$

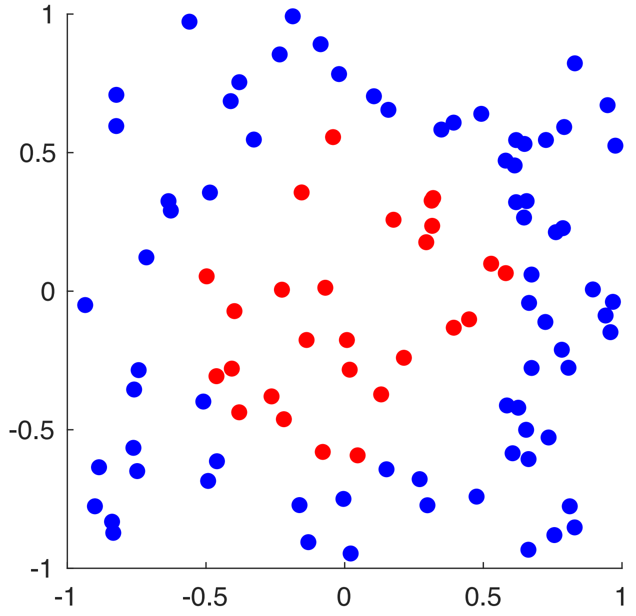
$$w_* = \sum_{i=1}^{m} \mu_{*,i} t_i x_i$$

$\mu_{*,i} > 0 \Rightarrow x_i$ *ist ein Stützvektor*

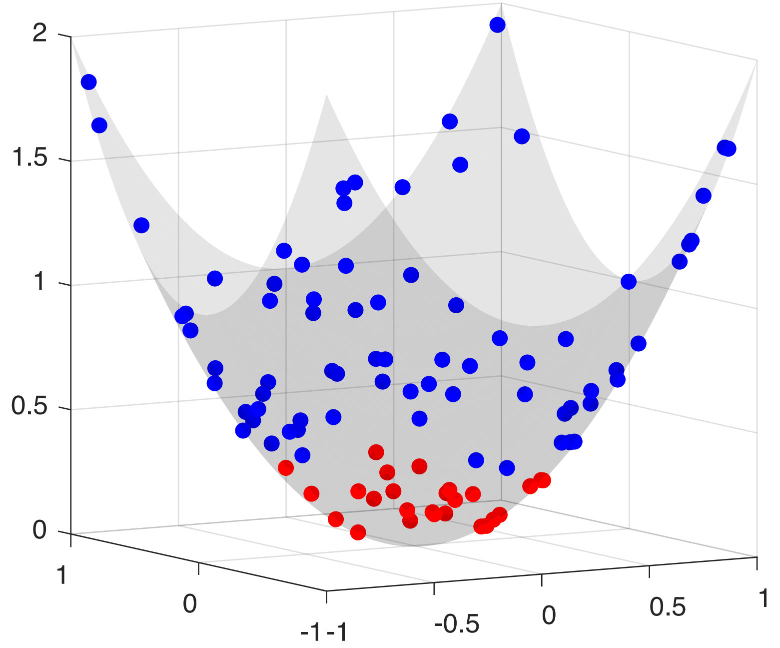
$$t(x) = \operatorname{sgn}\left(\sum_{i:x_i \text{ ist ein Stützvektor}} \mu_{*,i} t_i x_i^T x - \gamma_*\right)$$

Skalarprodukte mit Stützvektoren bestimmen die Klasse von x

Beispiel



$$\varphi\colon x \mapsto (x, z)$$
$$z = \|x\|_2^2$$



- Die Klassen von $x_1, \ldots, x_m \in \mathbb{R}^2$ sind nicht linear separierbar
- Setze $y_i = \varphi(x_i) \in \mathbb{R}^3$ mit $\varphi(x) = (x, \|x\|_2^2)$

Die Klassen von $y_1, \ldots, y_m$ sind linear separierbar!

Klassifizierungsregel

- Wir wenden die SVM auf $y_1, \ldots, y_m$ an $\Rightarrow$ Klassifizierungsregel

$$t(x) = \operatorname{sgn}\left(\sum_i \mu_{*,i} t_i \varphi(x_i)^T \varphi(x) - \gamma_*\right)$$

- φ tritt in die Klassifizierungsregel nur durch ihren Kernel ein:

$$K(u, v) = \varphi(u)^T \varphi(v)$$

- Für ein beliebiges $\varphi: \mathbb{R}^n \to \mathbb{R}^M$ hat das duale Problem dieselbe Anzahl der Variablen m und Nebenbedingungen $m + 1$

Das primale Problem hat M Variablen

$M = \infty$ *ist auch möglich!*

Kernel-Trick

- Durch eine geeignete Abbildung φ in einen hochdimensionalen Raum kann man die SVM auf nicht linear separierbare Daten anwenden
- φ erscheint in der Klassifizierungsregel durch ihren Kernel K

Man kann direkt nach einen geeigneten Kernel $K(u,v)$ suchen

- Satz von Mercer.

$$K(u,v) = \varphi(u)^T\varphi(v) \Leftrightarrow \begin{pmatrix} K(x_1,x_1) & \cdots & K(x_1,x_m) \\ \vdots & \ddots & \vdots \\ K(x_m,x_1) & \cdots & K(x_m,x_m) \end{pmatrix} \succcurlyeq 0$$

mit $\varphi\colon \mathbb{R}^n \to \mathbb{R}^M, M \leq \infty$

$$\forall x_1, \dots, x_m, \forall m = 1,2,\dots$$

Zusammenfassung

- Sattelpunkte und Dualität
- Dualität für restringierte Optimierungsprobleme
- Beispiel: Support Vektor Machines

Nächstes Video

- 6a. Lineare Programmierung: Geometrische Grundlagen