VORLESUNGSNOTIZEN

Optimierung

Wintersemester 24/25

Emma Bach

Vorlesung gehalten von
Prof. Rolf BACKOFEN
und
Prof. Moritz DIEHL

March 10, 2025

Inhalt

Chapter 1

Einführung

Ein Optimierungsproblem besteht aus einer **zulässigen Menge** G und einer **Zielfunktion** $f : G \to H$, wobei wir im Allgemeinen von $H = \mathbb{R}$ ausgehen werden. Wir schreiben dann zum Beispiel

$$\min_{x \in \mathbb{R}} f(x),$$

um das Problem "finde den kleinsten Wert, den $f(x)$ bei reellem x annimt" zu notieren, und

$$\operatorname*{argmin}_{x \in \mathbb{R}} f(x) = 4$$

für das Problem "finde die kleinste reelle Zahl x, sodass $f(x) = 4$ ist.

Historisch ist das Feld der mathematischen Optimierung unter Nebenbedingungen stark verankert im Feld der **Operations Research**, welches sich mit der Optimierung von Produktionskosten unter gegebenen Bedingungen beschäftigt. Heutzutage hat die Optimierung zahlreiche Anwendungen, z.B. in der Pfadplanung, in Computer Vision, in der Bioinformatik, im maschinellen Lernen oder im Hardwaredesign.

1.1 Anwendungsbeispiel: Support Vector Machines

Ein klassisches Problem des maschinellen Lernes ist die Klassifizierung von Daten durch eine lineare Entscheidungsgrenze - alle Punkte (x_i, y_i) über einer Ebene werden einer Klasse zugeordnet, und alle Punkte unter der Ebene einer anderen Klasse. Das Problem besteht daraus, eine Optimale Trennungsebene zu finden. Diese wird beschrieben durch eine Gleichung der Form

$$\hat{y}(x) = w^T x + w_0.$$

Sei w^* der Vektor $(w_0, w_1, \ldots, w_{n-1})$. Es stellt sich heraus, dass das zu lösende Optimierungsproblem gegeben ist durch:

$$\begin{array}{cc} \operatorname*{argmin}_{w^* \in \mathbb{R}^n} & \frac{1}{2}||{w^*}^2|| \\ \text{s.t.} & \forall i \; y_i \hat{y}(x_i) \geq 1 \end{array}$$

Wir wollen die Länge des Vektors w^* minimieren, da dies zu einem größeren Abstand zwischen unserer Ebene und den Datenpunkten führt, wodurch unser Modell besser generalisiert. Die Nebenbedingung entspricht der Anforderung, dass alle Punkte korrekt klassifiziert werden sollen.

1.2 Typische Problemklassen

Optimierungsprobleme können gemäß diverser Kriterien klassifiziert werden:

- Probleme mit Nebenbedingungen vs Probleme ohne Nebenbedingungen

- Optimierung mit Variablen aus verschiedenen Mengen, insbesondere kontinuierliche Variablen vs diskrete Variablen
- Lineare vs nichtlineare Funktionen
- Eindimensionale vs mehrdimensionale Funktionen
- Konvexe Funktionen vs nicht konvexe Funktionen
- Konvexe Mengen vs nicht konvexe Mengen

Diese verschiedenen Problemklassen führen zu unterschiedlich schwierigen Problemen. Insbesondere sind konvexe Probleme einfacher zu lösen als nicht konvexe Probleme, kontinuierliche Probleme sind in der Regel einfacher zu lösen als diskrete Probleme, und lineare Probleme sind einfacher als nichtlineare Probleme. Relevante Fragen sind dann:

- Wie schnell kovergiert das Verfahren zu einer Lösung? Wie viele Iterationen sind nötig? Was ist die Komplexität (in O-Notation) einer einzelnen Iteration?
- Konvergiert das Verfahren immer gegen ein Globales Optimum? Falls nein, gibt es garantierte obere/untere Schranken für die maximale Abweichung vom globalen Optimum?

1.3 Konvexität

Eine Menge G ist **konvex**, wenn für beliebige Punkte $x, y \in G$ auch beliebige lineare Interpolationen zwischen den Punkten in der Menge enthalten sind:

$$x, y \in G \implies \{(1-\alpha)x + \alpha y \mid \alpha \in [0,1]\} \subset G$$

Intuitiv entspricht das der Forderung, dass zwischen für alle Paare von Punkten eine Verbindungsstrecke zwischen den Punkten in der Menge enthalten sein muss.

Die **konvexe Hülle** einer Menge G ist die kleinste konvexe Menge H sodass $G \subset H$.

Analog zur Definition einer konvexen Menge ist eine **konvexe Funktion** definiert als eine Funktion, für die gilt:

$$x, y \in G \implies f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) \ \forall \alpha \in [0,1]$$

Dies entspricht der Forderung, dass jede Verbindungsstrecke zwischen zwei Punkten auf dem Graphen unterhalb des Graphen liegen muss.

Eine nicht konvexe Funktion, in der jedes lokale Minimum auch ein globales Minimum ist, wird als **quasikonvex** bezeichnet. Da dies dem Hauptvorteil von konvexen Funktionen in der Optimierung entspricht, ist die Optimierung von quasikonvexe Funktionen ebenfalls einfacher als die Optimierung allgemeiner nichtlinearer Funktionen.

Chapter 2

Gradientenverfahren