Inhaltsverzeichnis

Learning from Data

Mathematische Methoden des digitalen Zeitalters

1 Optimierung

1.1 Grundlagen der Optimierung

Bei der Optimierung betrachtet man vor allem Funktionen, welche Elemente eines n-dimensionalen Vektorraumes auf reelle Zahlen abbilden:

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Beispielhafte Anwendungszwecke wären die Minimierung oder Maximierung der Kosten bzw. der Einnahmen eines Unternehmens. Diese würden dem reellen Funktionswert der Funktion entsprechen und ergäben sich aus den Eigenschaften des Unternehmens wie Gehalt, Anzahl der Mitarbeiter, Marketing-Ausgaben usw., welche man zusammenfassend als Vektor eines n-dimensionalen Vektorraumes darstellen könnte.

1.2 Regression

Theorem 1 *(Methode der kleinsten Quadrate)* Beispiel für die Lösung eines solchen Minimierungsproblems wäre die Methode der kleinsten Quadrate. Gegeben ist hierbei eine lineare Funktion f und Punkte $P_i = (x_i, y_i)$, welche näherungsweise auf der Geraden liegen:

$$f(x) = ax$$
$$y_i = f(x_i) + \epsilon_i$$

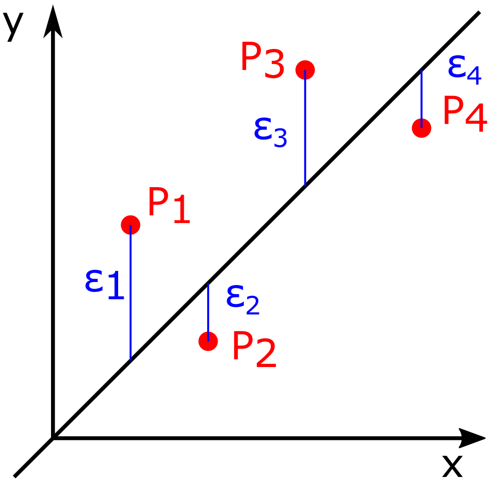

Abb. 1.1: Konvexe Funktion

Wenn nun der Abstand ϵ_i minimiert werden soll, dann gilt:

$$\min h(x) = \frac{\sum_{i=1}^{n} (y_i - ax_i)^2)}{e(y_i, x_i)}$$

Hierbei könnte $e^2, \mid e \mid, h(e)$ exemplarische Möglichkeiten seien. Zur Berechnung dieses Minimums gibt es nun verschiedene Möglichkeiten:

1. Gewöhnliche Tiefpunktberechnung
   z. B. $\quad h'(a) = 0 \quad \wedge \quad h''(a) > 0$
2. Gradientenabstiegsmethode
   $a^{t+1} = a^t - \lambda \cdot \delta f(a^t)$
   Man findet also ein $\lambda > 0$, sodass $f(a^{t+1}) < f(a^t) \quad \mid \delta f(a^{(t)}) \neq 0$
   $g^+(\lambda) = f(a^+ - \lambda \Delta f(a^t))$
   $\lambda^+ = argmin(g^+(\lambda))$

DEFINITION 1 *(Lokales und globales Minimum)* Für $f : D \to \mathbb{R}$ ist $x \in D$ ein lokales Minimum, wenn mindestens eine Umgebung N existiert, sodass
$\forall_{y \in N}$ gilt: $f(y) >= f(x)$
$x \in D$ ist dann ein globales Minimum, falls $N = D$.

1.3 Konvexität

DEFINITION 2 *(Konvexe Funktion)* Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist dann eine konvexe Funktion, falls $\forall_{x,y \in D}$ gilt: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (a-\lambda)f(y) \quad | \forall_{y \in [0,1]}$

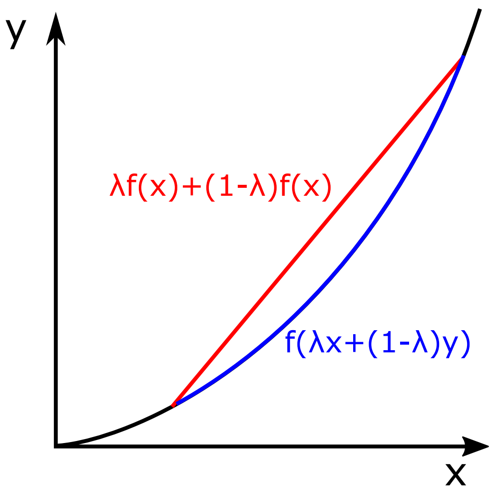

Abb. 1.2: konvexe Funktion

1.3.1 Beispiel zur Konvexität

$$\begin{aligned} f(u) &= \| u \|_2^2 \qquad | u \in \mathbb{R} \\ f(\lambda u + (1-\lambda)v) &= \| \lambda u + (1-\lambda)v \|_2 \\ &\leq \| \lambda u \|_2 + \| (1-\lambda)v \|_2 \\ &\leq \lambda \| u \|_2 + (1-\lambda) \| v \|_2 \end{aligned}$$

1.4 Konvexe Mengen

Im Rahmen der Optimierung von konvexen Funktionen ist es erforderlich, den Begriff der konvexen Menge einzuführen. Um diese Thematik anschaulich darzustellen, verwenden wir zunächst verschiedene geometrische Figuren in Abbildung 1.4, von denen einige konvex sowie andere wiederum nicht konvex sind.

Ist es möglich eine direkte Verbindungsstrecke zwischen zwei beliebigen Punkten zu finden, die selbst ebenfalls in der Menge liegt, so ist die Menge konvex. Formal bedeutet das:

DEFINITION 3 *(Konvexe Menge)* Eine Menge X heißt konvex, falls für alle $x, y \in X$ und $\lambda \in [0,1]$ gilt, dass $\lambda x + (1-\lambda) \in X$.

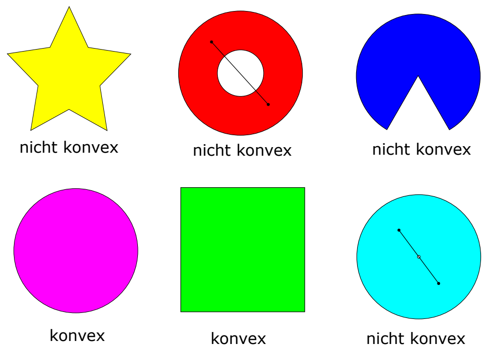

Abb. 1.3: Beispiele konvexer Mengen

1.5 Beispiele konvexer Mengen

Kugel: Ein Beispiel für eine konvexe Menge ist die Menge $S = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ der Vektoren, die eine Kugel mit dem Radius $r = 1$ und Mittelpunkt im Ursprung beschreibt.

Quadrant: Auch die Quadranten des kartesischen Koordinatensystems lassen sich durch eine konvexe Menge erfassen; Beispiel ist der erste Quadrant $L = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n\}$.

Eistüte: Die Eistüte bzw. der quadratischen Kegel $Q = \{(x,t) \mid x \in \mathbb{R}^{n-1}, t \in \mathbb{R}, \|x\|^2 \leq t\}$ ist ein weiteres Beispiel für eine konvexe Menge.

Box: Eine Box mit der Seitenlänge 2, deren Mittelpunkt im Ursprung liegt, ist eine konvexe Menge $B = \{x \in \mathbb{R}^n \mid |x_i| \leq 1, i = 1, \ldots, n\}$.

1.6 Konvexe Programme

Ziel eines konvexen Programmes ist die Optimierung einer Funktion f, die vom Parameter x abhängig ist. Hierbei sind sowohl die Zielfunktion als auch eventualle Nebenbedingungen konvex.

DEFINITION 4 *(Lösung konvexer Programme)* Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion und $X \subseteq \mathbb{R}^n$ konvex, so ist die Lösung des Programms

$$\begin{aligned} \text{minimiere} \quad & f(x) \\ \text{sodass} \quad & x \in X \end{aligned}$$

das globale Minimum der Funktion $f : X \to \mathbb{R}$.

Im Folgenden betrachten wir einige Beispiele.

Lineare Programme: Für lineare Programme gilt $f(x) = \langle c, x \rangle$ und $X = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Ein Beispiel für ein lineares Programm wäre

$$\begin{aligned} \text{minimiere} \quad & f(x) = x_1 \\ \text{sodass} \quad & x_1 \leq 1, x_2 \leq 1. \end{aligned}$$

Quadratische Programme: Im Fall quadratischer Programme ist die Zielfunktion eine quadratische Funktion $f(x) = x^{\mathrm{T}}Qx$. Ein mögliches Beispiel ist

$$\begin{aligned} \text{minimiere} \quad & f(x) = x_1^2 + x_2^2 \\ \text{sodass} \quad & x_1 \leq 1, x_1 + x_2 \leq 1. \end{aligned}$$

1.7 Gradientenmethode

Folgender Algorithmus implementiert die Gradientenmethode, die einen wichtigen Teil des Gebietes der Optimierung darstellt. Funktionieren tut sie wie folgt:

```
procedure calculateMinimum
F(x) = (x^2 - 2)^2
F'(x) = 4x^3-8x

x = 10.0
λ = 0.001
    for i = 1, ..., m do
        λ = λ + 0.001 · i
        for j = 1, ..., n do
            x = x - λ · F'(x)
        end for
    end for
Print x
end procedure
```

Erklärung $F(x)$ beschreibt die Funktion, deren globales Minimum wir finden wollen. $F'(x)$ ist dementsprechend die Ableitung der Funktion $F(x)$. Mit $x = 10.0$ setzen wir den Schätzwert, ab dem optimiert wird. λ bekommt einen niedrigen Wert deklariert, damit man sich in kleinen Schritten dem Minimum annähern kann. Dies wiederum geschieht in zwei For-Schleifen, deren Inhalt in diesem Fall je 8 mal durchlaufen wird. In der äußeren Schleife sorgen wir dafür, dass unser Lambda größere Werte annimmt. In der inneren For-Schleife wird die Schätz-Variable x mit Hilfe der Ableitung und λ angepasst. Daraus folgt:

$$f(x + \lambda \nabla f(x)) \leq f(x)$$