Inhaltsverzeichnis

Learning from Data

Mathematische Theorien und Methoden für das digitale Zeitalter

Titelbild fehlt
--
Bitte zusenden!

1 Aus der Kursbeschreibung

Philipp Moritz KL, Fanny Yang KL

Test.

Jeder Kurs beginnt mit einem Teil „1 Aus der Kursbeschreibung«, entnommen aus dem Programmheft. Hier sollte nur der den Inhalt des Kurses beschreibende Teil der Kursbeschreibung erscheinen, nicht die Erwartungen an die Teilnehmenden.

Gegen Ende des Kurses können die Teilnehmenden je nach Interesse weitere Themen diskutieren. Es kann zum Beispiel das ${H_2}^+$-Ion als ein

2 Vektoren

In der linearen Algebra nehmen Vektoren eine zentrale Rolle ein. Ein Vektor wird in der Schule als eine Menge an Pfeilen gelehrt, die parallel, gleichgerichtet und gleich lang sind. Betrachten wir jedoch nicht nur den dreidimensionalen Raum, sondern $\mathbb{R}^n$, so ist die Vorstellung eines Pfeils nicht immer möglich . Dies bedingt die Notwendigkeit einer anderen Definition.

Ein *Vektor* ist eine Aufzählung von Objekten und beschreibt eine Verschiebung. Eine solche Aufzählung entspricht der Definition eines *Tupels*. Entscheidend bei Tupeln ist die Reihenfolge der Objekte, die auch mehrfach vorkommen können.

$$\text{n-Tupel } (a_1, a_2, \ldots, a_n)$$

2.1 Vektor- & Unterraum

Vektoren bilden die Elemente eines *Vektorraumes V*. Addieren wir zwei Vektoren eines Vektorraumes oder multiplizieren wir sie mit einem Skalar, so ist die Summe bzw. das Produkt ebenfalls ein Element des Vektorraumes.

$$\begin{aligned} (1)\ u, v \in V,\ & u + v \in V \\ (2)\ u \in V, \lambda \in \mathbb{R},\ & \lambda u \in V \end{aligned}$$

Ein *Unterraum* ist eine Teilmenge eines Vektorraumes. Es gelten für sie die oben genannten Eigenschaften eines Vektorraumes.

$$\begin{aligned} U \subseteq V, & \text{wenn gilt:} \\ (1)\ u, v \in U,\ & u + v \in U \\ (2)\ u \in U, \lambda \in \mathbb{R},\ & \lambda u \in U \end{aligned}$$

2.2 Span

Alle möglichen Vektoren eines Vektorraumes werden durch den sogenannten *Span* dargestellt. Der Span ist die Menge aller möglichen Linearkombinationen der Basisvektoren.

$$\text{Span } (V_1, \ldots, V_k) = \{U \in V : U = \lambda_1 v_1 + \cdots + \lambda_k v_k\} \quad \text{mit } \lambda_1, \ldots, \lambda_k \in \mathbb{R}$$

2.3 Basis & Dimension

Die Menge der Basisvektoren wird *Basis* eines Vektorraumes genannt. Jeder Vektorraum besitzt eine *Dimension p*, die durch die Anzahl der Basisvektoren bestimmt wird.

Übung
Beweise, dass ein Span immer ein Vektorraum ist.
BEWEIS Wir nehmen an:

$$U \in Span(V_1, \dots, V_k)$$
$$V \in Span(V_1, \dots, V_k)$$

$$U = \lambda_1 V_1 + \dots + \lambda_k V_k$$
$$V = \alpha_1 V_1 + \dots + \alpha_k V_k$$

Additionsregelung bei Vektorräumen (s.(1))

$$\begin{aligned} W_1 &= U + V \\ &= \lambda_1 V_1 + \dots + \lambda_k V_k + \alpha_1 V_1 + \dots + \alpha_k V_k \\ &= (\lambda_1 \alpha_1) V_1 + \dots + (\lambda_k \alpha_k) V_k \end{aligned}$$

Multiplikationsregelung bei Vektorräumen (s.(2))

$$\begin{aligned} W_2 &= \lambda_1 V_1 + \dots + \lambda_k V_k \\ \lambda W_2 &= (\lambda \lambda_1) V_1 + \dots + (\lambda \lambda_k) V_k \end{aligned}$$

□

2.4 Lineare Unabhängigkeit

$v_1, \dots, v_k$ mit $v_i \in V$ sind *linear unabhängig*, falls $\lambda_i v_i + \dots + \lambda_k v_k = 0$ nur für $\lambda_i = \dots = \lambda_k = 0$ gilt.

2.5 Norm und Skalarprodukt

DEFINITION 1 $||v||$ ist eine Norm im Vektorraum V falls

1. a) $||v|| \geq 0$ für alle $v \in V$
   b) $||v|| = 0$ nur für $v = 0$
2. $||u + v|| \leq ||u|| + ||v||$ für alle $v \in V$
3. $||\lambda v|| = |\lambda| ||v||$ für alle $\lambda \in \mathbb{R}$ und $v \in V$

Eine Norm kann nie negativ sein, ist die Norm 0 bedeuted das, dass der Vektor der Nullvektor ist. Wenn man zwei Vektoren addiert, ist die Norm des resultierenden Vektors genau die Summe der Normen der beiden Ausgangsvektoren. Außerdem ist die Norm eines Vektors, der mit einem reelen Faktor multipliziert wurde gleich dem Produkt aus der Norm des Vektors und dem Betrag des Faktors.

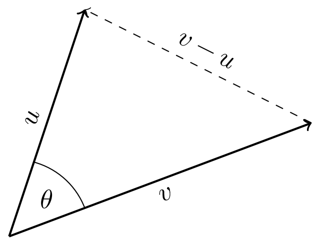
DEFINITION 2 $f : V \times V \to \mathbb{R}$ ist ein Skalarprodukt wenn

1. $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$ (Bilinearität)
2. $\langle u, v \rangle = \langle v, u \rangle$ für alle $u, v \in V$
3. $\langle v, v \rangle \geq 0$ mit $\langle v, v \rangle = 0$ nur für $v = 0$

Üblicherweise wird das Skalarprodukt mit der folgenden Formel berechnet.

$$\langle u, v \rangle = u_1 v_1 + \dots + u_n v_n$$

An der ausgeschriebenen Form lässt sich überprüfen, ob es sich per Definition dabei wirklich um ein Skalarprodukt handelt. In der ausgeschriebenen Form wird aus $\langle \alpha u + \beta v, w \rangle$ dabei $(\alpha u_1 + \beta v_1) w_1 + \dots + (\alpha u_n + \beta v_n) w_n$. Durch Ausmultiplizieren der Produkte erhalten wir die Form $\alpha u_1 w_1 + \beta v_1 w_1 + \dots + \alpha u_n w_n + \beta v_n w_v$. Wenn wir daraus nach zwei Summen sortieren und die Summanden mit α von denen mit β trennen können, wir die beiden Faktoren ausklammern und gelangen zu der aus der Definition geforderten Form $\alpha \langle u, w \rangle + \beta \langle v, w \rangle$.
Die zweite Bedingung können wir mit dem Kommutativgesetz beweisen, wenn wir dieses auf die ausgeschriebene Form anwenden. Für die dritte Bedingung ergibt sich $v_1 v_1 + \dots + v_n v_n = v_1^2 + \dots + v_n^2$. Da durch das Quadrieren keiner der Summanden kleiner als 0 werden kann und auch nur für $v_i = 0$ genau 0 werden kann, ist auch diese Bedingung erfüllt.



Da grundsätzlich $\langle v, v \rangle = ||v||^2$ gilt, kann über die Abbildung auch die Aussage getroffen werden, dass $||v - u||^2 = \langle v - u, v - u \rangle$. Mit der Bilinearität des Skalarproduktes und der eben getroffenen Aussage können wir die Gleichung zu

$$||v - u||^2 = ||v||^2 - 2 \langle v, u \rangle + ||u||^2$$

umformen. Zusätzlich folgt aus dem Kosinussatz, dass $||v - u||^2 = ||u||^2 + ||v||^2 - 2||u|| ||v|| \cos \theta$. Setzt man beide Terme gleich, gelangt man über einige wenige Umformungen zu der Form $\langle u, v \rangle = ||u|| ||v|| \cos \theta$, was ebenfalls eine Möglichkeit ist, das Skalarprodukt darzustellen.
Aus dieser Gleichung ergibt sich direkt die Cauchy-Schwarz Ungleichung. Da der Betrag des Kosinus sich nur zwischen 0 und 1 bewegt, gilt $\langle u, v \rangle \leq ||u|| ||v||$.
Eine häufig genutzte Norm ist die Euklidische Norm, welche mit $||v||_2 = \sqrt{v_1^2 + \dots + v_n^2}$ definiert ist. Mit den zuvor getroffenen Aussagen können wir beweisen, dass es sich bei der Euklidischen Norm wirklich um eine Norm handelt.
Die erste und die dritte Bedingung aus der Definition für eine Norm zeigen sich aus der ausgeschriebenen Form der Euklidischen Norm. Da alle

Komponenten dabei quadriert werden und die Wurzel gezogen wird muss die erste Bedingung stimmen. Ein Faktor λ, der ebenfalls quadriert in der Wurzel steht kann aus der Summe ausgeklammert werden und als Faktor vor die Wurzel gelangen. Der Betrag ergibt sich dabei daraus, dass λ nach dem Vorgang nur positiv sein kann, auch wenn es zuvor negativ war.
Um zu beweisen, dass $\|u+v\| \leq \|u\| + \|v\|$ formen wir diese Ungleichung zu der bekannten Cauchy-Schwarz Ungleichung um. Hierzu quadrieren wir zunächst beide Seiten, ersetzen die quadrierten Normen durch Skalarprodukte und erhalten

$$|\langle u+v, u+v\rangle| \leq \langle u,u\rangle + 2\|u\|\|v\| + \langle v,v\rangle$$

Die linke Seite der Ungleichung kann aufgrund der Bilinearität als $|\langle u,u\rangle| + 2|\langle u,v\rangle| + |\langle v,v\rangle|$ geschrieben werden. Dadurch können wir auf beiden Seiten $|\langle u,u\rangle|$ und $|\langle v,v\rangle|$ subtrahieren und kommen so auf $\|u+v\| \leq \|u\| + \|v\|$, womit bewiesen ist, dass die Euklidische Norm eine Norm ist.

3 Lineare Abbildungen

In diesem Abschnitt werden wir die elementarste Form einer Funktion kennenlernen, die einen Vektorraum auf einen anderen Vektorraum abbildet. Diese Funktionen nennt man *lineare Abbildungen*.

DEFINITION 3 Eine Funktion $f : U \rightarrow V$ ist linear, wenn folgende Bedingungen gelten:
1.
$$f(u+v) = f(u) + f(v) \forall\ u, v \in U$$

2.
$$f(\lambda v) = \lambda f(v)\ \forall\ v \in U\ und\ \lambda \in \mathbb{R}$$

Das besondere an solchen Abbildungen ist, dass sie von *Matrizen* der Form $A \in \mathbb{R}^{n \times m}$ definiert werden, wobei n die Dimension von U und m die Dimension von V ist.

3.1 Matrix

Eine Matrix ist eine spezielle Anordnung von $n \cdot m$ Zahlen in n Zeilen und m Spalten. Man kann eine Matrix aber auch als eine Zusammenfassung von m Vektoren mit n Komponenten ansehen, dabei bildet ein Vekotor eine Spalte. Die Komponente in der i-ten Zeile und j-ten Spalte wird als a_{ij} geschrieben und die ganze Spalte mit a_j notiert.

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1m} \\ a_{21} & a_{22} & ... & a_{2m} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nm} \end{pmatrix} = (a_1, a_2, ..., a_m)$$

Da wir nun wissen was eine Matrix ist, führen wir zwei neue Operationen ein.
1. Matrix-Vektor-Multiplikation
2. Matrix-Matrix-Multiplikation

3.2 Matrix-Vektor-Multiplikation

Bei einer Matrix-Vektor-Multiplikation wird eine Matrix $A \in \mathbb{R}^{n \times m}$ mit einem Vektor $v \in \mathbb{R}^m$ multipliziert. Dabei wird jede Zeile der Matrix mit dem Vektor skalarmultipliziert; deshalb ist es wichtig, dass die Matrix so viele Spalten wie der Vektor Dimensionen hat. Eine Multiplikation zwischen Matrizen und Vektoren, bei denen die Anzahl nicht übereinstimmt, ist nicht definiert.

$$\begin{aligned} Av &= \begin{pmatrix} a_{11} & a_{12} & ... & a_{1m} \\ a_{21} & a_{22} & ... & a_{2m} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nm} \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ ... \\ v_m \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \times v_1 + a_{12} \times v_2 + ... + a_{1m} \times v_m \\ a_{21} \times v_1 + a_{22} \times v_2 + ... + a_{2m} \times v_m \\ ... \\ a_{n1} \times v_1 + a_{n2} \times v_2 + ... + a_{nm} \times v_m \end{pmatrix} \end{aligned}$$

2.1: Matrix mit Vektor multiplizieren

Allgemein lässt sich für die Matrix-Vektor-Multiplikation sagen

$$(Av)_i = \sum_{j=1}^{n} a_{ij} v_j$$

3.3 Matrix-Matrix-Multiplikation

Eine Matrix-Matrix-Multiplikation ist nichts anderes als mehrere Matrix-Vektor-Multiplikationen hintereinander. Haben wir zwei Matrizen $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times p}$ und berechnen AB nehmen wir jede Spalte von B als Vektor und multiplizieren ihn mit A wie oben beschrieben. Die Vektoren die man als Ergebnisse erhält, fasst man wieder in einer Matrix zusammen mit der Dimension $\mathbb{R}^{n \times p}$. Allgemein können wir sagen:

$$(AB)_{il} = \sum_{n=i}^{m} a_{ij} \times b_{jl}$$

Übung
Beweise, dass eine Komposition $h(x) = f \circ g(x)$ aus den linearen Abbildungen f und g, auch eine lineare Abbildung ist.
BEWEIS Da f und g linear sind gelten die Bedin-

gungen aus Definition 3:

$$\begin{aligned} h(u+v) &= f \circ g(u+v) \\ &= f(g(u+v)) \\ &= f(g(u)+g(v)) \\ &= f(g(u))+f(g(v)) \\ &= f \circ g(u) + f \circ g(v) = h(u)+h(v) \end{aligned}$$

Des Weiteren gilt:

$$\begin{aligned} h(\lambda v) &= f \circ g(\lambda v) \\ &= f(g(\lambda v)) \\ &= f(\lambda g(v)) \\ &= \lambda f(g(v)) \\ &= \lambda f \circ g(v) = \lambda h(v) \end{aligned}$$

Da beide Bedingungen aus Definition 3 für h erfüllt sind, ist auch eine Komposition aus zwei linearen Abbildungen linear. □

4 Analysis

Analysis ist ein wichtiges Teilgebiet der Mathematik. Sie ist die Grundlage für Optimierung und somit auch von großer Bedeutung beim Maschinellen Lernen.
Im Kurs war für uns besonders die Stetigkeit und Differenzierbarkeit von ein- und mehrdimensionalen Funktionen wichtig.
Um uns Klarheit üuber diesen Bereich der Mathematik zu verschaffen, lösen wir - unter anderem - die folgenden Aufgaben.

Aufgabe 1:
Aufgabenstellung: Beweisen Sie, dass eine Funktion $f : \mathbb{R} \to \mathbb{R}$, welche differenzierbar an der Stelle $x \in R$ ist auch stetig an der Stelle $x \in R$ ist.

Lösung:
Zuerst überlegen wir, welche Bedingungen bereits in der Aufgabenstellung gegeben sind.
Eine Funktion ist an der Stelle x differenzierbar, wenn gilt:

$$f(x+h) = f(x) + l_x(h) + r(h) \text{ mit } \lim_{h\to 0} \frac{r(h)}{h} = 0 \tag{2.1}$$

Diese Funktion wäre außerdem stetig, wenn $\lim\limits_{w\to x} f(w) = f(x)$.
Da die Funktion $f(x+h)$ äquivalent zur Funktion $f(w)$ sein soll, ersetzen wir in der Bedingung für die Stetigkeit $f(w)$ mit der gesamten Funktion von $f(x+h)$. Somit gilt:

$$\lim_{h\to 0} f(x) + l_x(h) + r(h) = f(x).$$

Damit die Bedingung der Stetigkeit erfüllt ist, muss der Grenzwert dieser Funktion gleich dem Funktionswert sein. Um diese Bedingung zu erfüllen, müssen wir beweisen, dass die Teile der Funktion $l_x(h)$ und $r(h)$ gegen Null gehen. Wir beginnen mit dem Teil $r(h)$. Würde $r(h)$ nicht gegen Null gehen, so würde $\frac{r(h)}{h}$ nicht gegen Null gehen. Die Funktion $r(h)$ muss gegen Null gehen, da die Bedingung der Differenzierbarkeit gilt. Jetzt müssen wir zeigen, dass $l_x(h)$ auch gegen Null geht. Dieser Teil der Funktion kann auch als $f'(x) \cdot h$ dargestellt werden. Da wir annehmen, dass h gegen Null geht und somit ein Faktor von $l_x(h)$ Null ist, wird die Funktion Null. Damit ist bewiesen, dass $\lim\limits_{h\to 0} f(x) = f(x)$.

Aufgabe 3:
Bei der nachfolgenden Aufgabe soll bewiesen werden, dass es immer möglich ist für eine differenzierbare Funktion f einen Skalar λ zu finden, damit folgende Aussage gilt:

$(1) f(x + \lambda \nabla f(x)) \leq f(x)$

Manche von Ihnen werden sich jetzt fragen, was das Zeichen ∇ bedeutet. Es steht für den Gradienten, welcher im Grunde die Ableitung einer Funktion darstellt. Dadurch lässt sich die Ungleichung (1) in folgende Form umwandeln:

$f(x) + Df(x)h + r(h) \leq f(x)$

Hierbei ist h = $\lambda\nabla$ f(x) und r(h) das Restglied .
Das Ganze kann man noch weiter umschreiben, indem man das Skalarprodukt bildet:

$f(x) + \lambda < \nabla f(x), \nabla f(x) > + r(h) \leq f(x)$

Anschließend kann man von der Regel Gebrauch machen, dass sich ein Skalarprodukt, das in der Form $< u, u >$ vorliegt, in die Ausprägung der Euklidischen Norm von $||u||^2$ umgewandelt werden kann.
Daraus folgt:

$f(x) + \lambda \cdot ||\nabla f(x)||^2 + r(h) \leq f(x)$

Um nun die obige Ungleichung (1) zu beweisen, muss man zeigen, dass
$\lambda||\nabla f(x)||^2 + r(h) \leq 0$ ist.
Dafür wird λ ausgeklammert. Anschließend erhält man:

Da $\|\nabla f(x)\|$ quadriert wird, wird dieser Term automatisch positiv. Da man den gesamten Term aber negativ haben will, wählt man $\lambda < 0$. Schließlich kennt jeder die Regel, dass ein Produkt negativ wird, falls eine ungerade Zahl von Faktoren negativ ist. Da wir λ negativ gewählt haben, muss $\|\nabla f(x)\|^2(2)\lambda(\|\nabla f(x)\|^2 + \frac{r(h)}{\lambda}$ positiv sein.
Bei $\|\nabla f(x)\|^2\|$ stellt dies aus den eben genannten Gründen kein Problem dar. Deswegen muss nur noch gezeigt werden, dass es mindestens einen Wert von $r(h)$ gibt, der größer oder gleich 0 ist. Ansonsten würde der Wert in der Klammer negativ werden können und damit die unsprüngliche Aussage (1) widerlegen.
Damit wir zeigen können, dass $r(h) >= 0$ ist, formten wir zuerst den Term (2) um:

$$\lambda(\|\nabla f(x)\|^2 + \frac{r(h)}{\lambda})$$

$$\Leftrightarrow \lambda(\|\nabla f(x)\|^2 + \frac{r(\lambda\nabla f(x))}{\lambda})$$

Bei dieser Umformung wurde h wieder mit $\lambda\nabla f(x)$ ersetzt.
$\leftrightarrow \lambda\|\nabla f(x)\| \cdot (\|\nabla f(x)\| + \frac{r(\lambda\nabla f(x))}{\lambda\|f(x)\|})$
Durch diese Umformungen schlussfolgerten wir, dass für die Funktion $h(x)$ folgendes gelten muss:
(3) $h(\lambda) = \|\nabla f(x)\| + \frac{r(\lambda\nabla f(x))}{\lambda\|\nabla f(x)\|} >= 0$

Da in $\|\nabla f(x)$ kein λ steht ist dieser Term konstant. Gleichzeitig wird durch das Bilden der euklidischen Norm festgelegt, dass ein Wert größe/gleich null vorliegt.
Deswegen muss nur noch gezeigt werden, dass dder zweite Summand ebenfalls positiv werden kann. Dazu bildeten wir den Grenzwert.
$\lim\limits_{h \to 0} \frac{r(\lambda\nabla f(x))}{\lambda\|\nabla f(x)\|} = \frac{r(\lambda\nabla f(x))}{\|\lambda\nabla f(x)\|}$

Man kann $\lambda\nabla f(x)$ wieder durch h ersetzen, woraus folgt:
$\lim\limits_{h \to 0} \frac{r(h)}{h}$

Der Grenzwert des Restbetrages strebt gegen null. Daraus lässt sich schließen, dass es einen Punkt gibt, ab welchem die Ungleichung der Funktion $h(\lambda)$ (3) erfüllt ist und damit auch die anfängliche Aussage (1).

Aufgabe 6:

In der letzten Aufgabe soll die Produktregel bewiesen werden.
Dafür waren zwei Funktionen gegeben: Einmal die Funktion f, die vom $\mathbb{R}^m$ in den $\mathbb{R}$ abbildet und die Funktion g, die ebenfalls vom $\mathbb{R}^m$ in den $\mathbb{R}$ abgebildet wird.

Es sind also die Funktion f(x) mit der Ableitung $f(x+h) = f(x) \cdot Df(x) \cdot h \cdot r_f(h)$ und die Funktion g(x) mit der Ableitung $g(x+h) = f(x) \cdot Dg(x) \cdot h \cdot r_g(h)$ gegeben.
Die Produktregel besagt, dass für die Funktion $h(x) = f(x) \cdot g(x)$ die Ableitung $Dh(x) = f(x) \cdot Dg(x) + Df(x) \cdot g(x)$ ist.
Damit dies bewiesen werden kann schreiben wir zuerst die Ableitung von $h(x+v)$ als Kombination aus den Ableitungen von $f(x)$ und $g(x)$ auf.

$$\begin{aligned} h(x+v) = (&f(x) + Df(x) \cdot v \\ &+ r_g(h)) \cdot (g(x) \\ &+ Dg(x) \cdot v + r_g(v) \end{aligned} \tag{2.2}$$

Dieser Term wird ausmultipliziert und wir erhalten:

$$\begin{aligned} h(x+v) = &f(x) \cdot g(x) \\ &+ f(x) \cdot D_g(x) \cdot v \\ &+ f(x) \cdot r_g(v) \\ &+ Df(x) \cdot v \cdot g(x) \\ &+ Df(x) \cdot v \cdot D_g(x) \cdot v \\ &+ Df(x) \cdot v \cdot r_g(v) \\ &+ r_g(v) \cdot g(x) \\ &+ r_g(x) \cdot Dg(x) \cdot v \\ &+ r_g(v) \cdot r_g(v) \end{aligned} \tag{2.3}$$

Von den ganzen Summanden müsssen letztendlich alle eliminiert werden, bis auf

$$f(x) \cdot g(x) + f(x) \cdot Dg(x) \cdot v + Df(x) \cdot v \cdot g(x) \tag{2.4}$$

Dafür müssen wir zeigen, dass die übrigen Summanden gegen Null gehen. Um das zu erreichen bilden wir den Grenzwert:

Damit die Summanden einzeln äufgesplittet"werden können, wenden wir die Dreiecksungleichung an:

Nun können wir von jedem Summanden einzeln den Limes bilden, um zu zeigen, dass der gesamte Grenzwert gegen Null geht. Dies ist aber nur der Fall, falls der Limes jedes Summanden gegen Null geht. Wir zeigen das exemplarisch für

$$\lim_{v \to 0} \frac{\|Df(x) \cdot r_g(v)\|}{\|v\|} \tag{2.5}$$

Df(x) ist eine Matrix mit einer Zeile und m Spalten in der jeweils c steht. Daraus folgern wir:

$$\lim_{v \to 0} \frac{m \cdot c \cdot \|v\| \cdot \|r_g(v)\|}{\|v\|} \tag{2.6}$$

$$m \cdot c \cdot \lim_{v \to 0} \frac{\|v\| \cdot \|r_g(v)\|}{\|v\| = 0} \tag{2.7}$$

Damit haben wir gezeigt, dass der Summand gegen Null geht. Das Selbe lässt sich auch mit den anderen Summanden zeigen und auch diese gehen alle gegen Null.
Die Produktregel ist damit bewiesen.

5 Wahrscheinlichkeitstheorie

1933 veröffentlichte der russische Mathematiker Andrey Kolmogorov sein Buch *Foundations of the Theory of Probability*, in dem er die drei Axiome der Wahrscheinlichkeitstheorie aufstellte. Es existiert ein Wahrscheinlichkeitsraum, der sich aus den drei Elementen $\Omega, \mathcal{F}, \mathbb{P}$ zusammensetzt.

1. Ω ist eine Menge mit endlicher Anzahl an Elementen. $|\Omega| = n$
2. $\mathcal{F}$ ist eine Menge der Ereignisse ď und eine Teilmenge von Ω. $\mathcal{F} \subseteq \Omega$
3. $\mathbb{P}$ bezeichnet die Sicherheit der Wahrscheinlichkeit, das Wahrscheinlichkeitsmaß. Dieses weist jedem Ereignis in $\mathcal{F}$ eine reelle Zahl zu. $\mathbb{P} : \mathcal{F} \longrightarrow \mathbb{R}$

— Kolmogorovs Axiome lauten:

1. Die Wahrscheinlichkeit eines Ereignisses ist eine positive, reelle Zahl.
   $\mathbb{P}(\mathbb{E}) \geq 0$ für alle $\mathbb{E} \in \mathcal{F}$.
2. Die Menge aller Ergebnisse bezeichnet man als sicheres Ereignis, das die Wahrscheinlichkeit 1 hat. $\mathbb{P}(\Omega) = 1$
3. Die Wahrscheinlichkeit der Vereinigung von disjunkten Ereignisssen ist gleich der Summe der Wahrscheinlichkeiten der disjunkten Ereignisse.

$$\mathbb{P}(\cup_{i=1}^{m} \mathbb{E}_i) = \sum_{i=1}^{n} \mathbb{P}(\mathbb{E}_i)$$

für alle $\mathbb{E}_1, ..., \mathbb{E}_m \in \mathcal{F}$

$\mathbb{E}_1, ..., \mathbb{E}_m$ ist disjunkt, falls $\mathbb{E}_i \cap \mathbb{E}_j = \emptyset$ für alle $i, j = 1, 2, ..., m$. Hierbei bezeichnet $\emptyset$ ein unmögliches Ereignis $\emptyset \in \mathcal{F}$.
Im Gegensatz dazu gibt es das sichere Ereignis mit $\Omega \in \mathcal{F}$.

Die Wahrscheinlichkeit wird errechnet durch $\mathbb{P} = \frac{|\mathbb{E}|}{|\Omega|}$. Alternativ definiert man:

$p_i \geq 0$ für $i \in \Omega$, sodass $\sum_{i \in \Omega} p_i = 1$, $p_i = \mathbb{P}(i)$

— Konsequenzen der Axiome:

1. $\mathbb{P}(\emptyset) = 0$
2. Monotonie: $A \subseteq B$ daraus folgt: $\mathbb{P}(A) \leq \mathbb{P}(B)$
3. $0 \leq \mathbb{P}(E) \leq 1$
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

— Zufallsvariablen

In einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ beschreibt die Funktion $X : \Omega \longrightarrow \mathbb{R}^k$ einen k-dimensionalen Zufallsvektoren. Im Fall von $k = 1$ redet man von einer Zufallsvariablen.

— Erwartungswert

Der Erwartungswert einer Zufallsvariablen beschreibt die Zahl, die die Zufallsvariable im Mittel annimmt.

$$\mathbb{E}[X] = \sum_{x \in \Omega} X(i) * \mathbb{P}(i) = \sum_{x \in \mathcal{X}} X * \mathbb{P}(X = x) wobei$$

X ein Wertebereich von x ist

— Varianz

Die Varianz kennzeichnet die Ausdehnung einer Wahrscheinlichkeit. Die Varianz einer Zufallsvariable X.

$$X = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

— Unabhängigkeit von Ereignissen

1. Wenn für Ereignisse $E, F \in \mathcal{F}$ gilt $\mathbb{P}(E \cap F) = \mathbb{P}(E) * \mathbb{P}(F)$ dann sind E und F unabhängig.
2. Wenn für die Zufallsvariablen

   $$X_1 : \Omega_1 \longrightarrow \mathbb{R}, \mathcal{F}_1, X_2 : \Omega_2 \longrightarrow \mathbb{R}, \mathcal{F}_2$$

   gilt

   $$\mathbb{P}(X_1 \in A \cap X_2 \in B) = \mathbb{P}(X_1 \in A * \mathbb{P} X_2 \in B)$$

   für alle $A, B \subseteq \mathbb{R}$

Marginal bestimmte Wahrscheinlichkeiten

$$\mathbb{P}(X_1 \in A) = \sum_{b \in \Omega_2} \mathbb{P} X_1 \in A \cap X_2 \in b)$$

Bedingte Wahrscheinlichkeit

$$\mathbb{P}(X_1 \in A | X_2 \in B) = \frac{\mathbb{P}(X_1 \in A \cap X_2 \in B)}{\mathbb{P}(X_2 \in B)}$$

Summe von Zufallsvariablen

$$X_1 : \Omega \longrightarrow \mathbb{R}, X_2 : \Omega \longrightarrow \mathbb{R}$$
$$Y = X_1 + X_2$$

Annahme: X_1 und X_2 sind unabhängig.

$$Y = l = \cup_{i=-\infty}^{\infty} X_1 = 1 \cap X_2 = l - i$$

$$\mathbb{P}(Y = l) = \sum_{i=-\infty}^{\infty} \mathbb{P}(X_1 = i \cap X_2 = l - i) = \sum_{i=-} \mathbb{P}(X_1 = i) * \mathbb{P}(X_2 = l - i)$$

Theorem: Liniarität des Erwartungswerts

Angenommen, wir hätten eine Funktion
$f : \mathbb{R} \longrightarrow \mathbb{R}$, für die gilt:

$$\int_{\mathbb{R}} f(x) dx = 1$$

dann könnten wir eine Wahrscheinlichkeitsverteilung folgendermaßen definieren:

$$\mathbb{P}(X \in A) = \int_{X \in A} f(x) dx$$