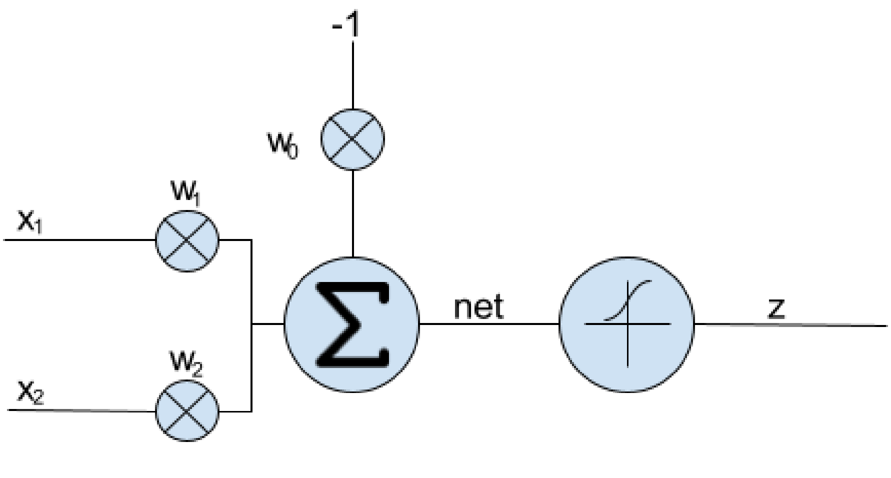
Die Mathematik der neuronalen Netze

Eine Einführung

By

Luca Ritz



2021
Luca Ritz

Inhaltsverzeichnis

Kapitel 1

Einführung

Kapitel 2

Die Mathematik der neuronalen Netze

2.1 Das Perceptron

2.1.1 Lernverfahren

2.1.2 Das Problem mit XOR und nichtlinearen Funktionen

2.2 Neuronale Netze

2.2.1 Lernverfahren mit Gradientenabstieg

2.2.2 Lernverfahren mit Backpropagation

2.2.3 XOR und die Lösung

Abbildungsverzeichnis

Glossar

Kapitel 3

Anhang

3.1 Die Ableitung 1. Grades

Die Ableitung 1. Grades beschreibt die Steigung an einem bestimmten Punkt der Funktion. In der Abbildung 3.1 wird eine Funktion $f(x)$ (in grün) gegeben. Die Steigung $f'(x)$ an einem bestimmten Punkt x ist rot markiert. Die Ableitung selbst ist wiederum eine Funktion und kann über diverse Ableitungsregeln aufgrund der gegebenen Funktion selbst gebildet werden.

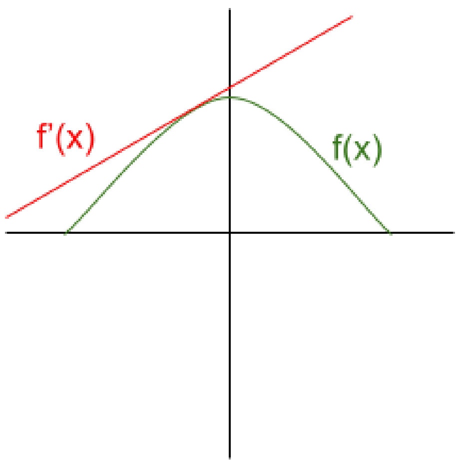


Abbildung 3.1: Steigung an einem bestimmten Punkt der Funktion $f(x)$

Die Ableitungsregeln, welche im Zuge der Erklärung des Lernprozesses eines neuronalen Netzwerks benötigt werden, sind nachfolgend ersichtlich.

Potenzregel $f(x) = x^n \longrightarrow f'(x) = n \cdot x^{n-1}$

Kettenregel $f(x) = u(v(x)) \longrightarrow f'(x) = u'(v(x)) \cdot v'(x)$

Es existieren viele weitere Ableitungsregeln, auf die hier nicht weiter eingegangen wird. Die Schreibweise der Ableitung einer Funktion nach einer Variablen lautet $f'(x) = \frac{\delta f(x)}{\delta x}$.

3.2 Die partielle Ableitung

Ist der Input einer Funktion mehrdimensional, das heisst, die Funktion f ist abhängig von mehreren Variablen, dann kann die Ableitung jeweils lediglich nach einer Variablen gebildet werden. Die übrigen Variablen werden als konstant angesehen. In dem Fall beschreibt die partielle Ableitung die Steigung an einem bestimmten Punkt der abgeleiteten Dimension. Es sei als Beispiel die Funktion $f(x,z) = x^2 + z^2 + 10$ gegeben. Diese wird nun partiell nach x sowie nach z abgeleitet.

$$f^x(x,z) = 2x \tag{3.1}$$

$$f^z(x,z) = 2z \tag{3.2}$$

Die hierbei angewendete Ableitungsregel ist die Potenzregel, welche bereits im vorangegangenen Kapitel erwähnt wurde.

3.3 Die Sigmoide

Als Aktivierungsfunktion wird die Sigmoidfunktion benutzt. Diese wird heutzutage meist nicht mehr eingesetzt aufgrund des schlechten Lernverhaltens in einigen Bereichen der Funktion. Die erwähnte Eigenschaft wird bei der Betrachtung der Ableitung ersichtlich.

$$sig(t) = \frac{1}{1+e^{-t}} \tag{3.3}$$

Geometrisch lässt sich die Funktion so interpretieren, dass eine gewisse Schwelle existiert, ab der die Funktion den Eingabewert auf eine 1 abbildet, in dem Jargon der Neuronen also „feuert“. Das Resultat lautet also entweder 0 oder 1. TODO: Bild

Für die weitere Verwendung ist nun vor allem die Ableitung der Sigmoide interessant, welche in den nachfolgenden Zeilen behandelt wird. Zuerst wird der Bruch durch eine andere Schreibweise (Exponent -1) dargestellt.

$$sig(t) = (1+e^{-t})^{-1} \longrightarrow y = 1+e^{-t}, f(t) = y^{-1} \tag{3.4}$$

Es handelt sich also um eine äussere und innere Funktion, wobei nun die Ableitungsregel 3.1 angewendet wird. Für die innere Ableitung wird nun noch die Regel für e verwendet. TODO

$$\frac{\delta f(t)}{\delta t} = (-1)\cdot(y(t))^{-2} \tag{3.5}$$

$$\frac{\delta y(t)}{\delta t} = (e^{-t})\cdot(-1) \tag{3.6}$$

Nach der Kettenregel resultiert:

$$\frac{\delta sig(t)}{\delta t} = (-1)\cdot(1+e^{-t})^{-2}\cdot(e^{-}t)\cdot(-1) \tag{3.7}$$

$$\frac{\delta sig(t)}{\delta t} = \frac{e^{-t}}{(1+e^{-t})^2} \tag{3.8}$$

Nun wird $\frac{1}{1+e^{-t}}$ ausgeklammert.

$$\frac{\delta sig(t)}{\delta t} = \frac{1}{1+e^{-t}}\cdot\frac{e^{-t}}{(1+e^{-t})} \tag{3.9}$$

Beim Zähler wird 1 dazuaddiert und abgezogen, damit der Faktor umgeformt werden kann.

$$\frac{\delta sig(t)}{\delta t} = \frac{1}{1+e^{-t}} \cdot \frac{e^{-t}+1-1}{(1+e^{-t})} \tag{3.10}$$

$$\frac{\delta sig(t)}{\delta t} = \frac{1}{1+e^{-t}} \cdot (\frac{1+e^{-t}}{(1+e^{-t})} - \frac{1}{1+e^{-t}}) \tag{3.11}$$

Es resultiert die Ableitung in bekannter Form 3.12

$$\frac{\delta sig(t)}{\delta t} = sig(t) \cdot (1 - sig(t)) \tag{3.12}$$

3.4 Der Gradient

Der Gradient beschreibt einen Vektor, welcher in Richtung des steilsten Anstiegs einer Funktion zeigt. Die Komponenten des Gradientenvektors bestehen aus den partiellen Ableitungen der Funktion an der jeweiligen Variablen.

$$\nabla f(x,y) \longrightarrow \begin{pmatrix} \frac{\delta f(x,y)}{\delta x} & \frac{\delta f(x,y)}{\delta y} \end{pmatrix} \tag{3.13}$$

Geometrisch kann dies an der Position $(x = 1, y = 1)$ für die Funktion $f(x,y) = x^2 + y^2$ wie in Abbildung 3.2 aussehen. Zu beachten sei hier, dass der eigentliche Gradient in der XY-Ebene liegt (blau). Der schwarze Vektor soll lediglich anzeigen, was eine Verschiebung in dieser Richtung bei der Eingabe der Variablen für den Ausgabewert der Funktion bedeutet.

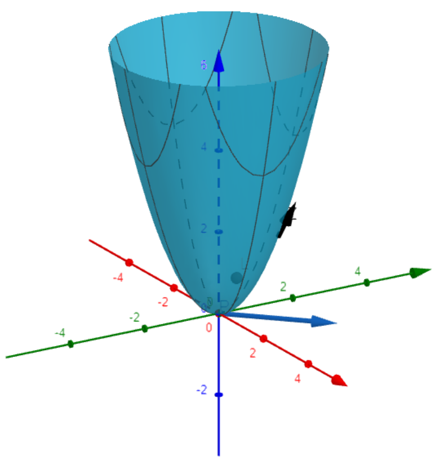


Abbildung 3.2: Gradient an der Position $(1,1)$

3.5 Das Gradientenabstiegsverfahren