

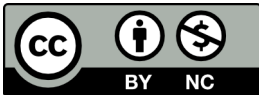
Deep Learning

DVA-Seminar 2018

Michael Schwab
Fakultät für Informatik
Hochschule für angewandte Wissenschaften Augsburg
Augsburg, Deutschland
michael.schwab@hs-augsburg.de

Zusammenfassung—

LIZENS



Deep Learning von Michael Schwab ist lizenziert unter einer Creative Commons Namensnennung-Nicht kommerziell 4.0 International Lizenz.

I. EINLEITUNG

II. KONVENTIONEN

KNN Künstliches Neuronales Netz
DL Deep Learning
IDS Iris-Datensatz
ML Machine Learning
IT Informationstechnik
CSV Coma Separated Values

III. GRUNDLAGEN

A. Datensätze

Eine essentielle Komponente des Deep Learning (DL) sind Daten. Ein Datensatz ist eine Sammlung von Daten in einem gleichen oder ähnlichen Format. Möchte man DL erlernen, steht man oft vor dem Problem keine passenden Daten parat zu haben. Aus diesem Grund gibt es verschiedene bereits gut klassifizierte Datensätze im Internet zu finden, die den Einstieg in das DL erleichtern. Ein Beispiel für einen dieser Datensätze ist der Iris-Datensatz (IDS)¹.

Der IDS wurde 1936 von dem Statistiker und Biologen Richard Fischer vorgestellt. Der Datensatz besteht aus 150 Einträgen und umfasst drei verschiedene Gattungen der Iris Blume. Jeder Eintrag einer einzelnen Blume wird jeweils durch sowohl durch die Blütenkelch Länge und Breite als auch durch die Blütenblatt Länge und Breite beschrieben. Die drei verschiedenen Gattungen der Iris Blume heißen Iris setosa, Iris virginica und Iris versitosa [vgl. 1].

Tabelle I zeigt einen Auszug aus dem IDS. Anhand diesem Beispiels sollen drei Begriffe erklärt werden, die

Tabelle I
EINTRÄGE AUS DEM IDS

	Feature 1	Feature 2	Feature 3	Feature 4	Label
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
140	6.7	3.1	5.6	2.4	Iris-virginica
141	6.9	3.1	5.1	2.3	Iris-virginica
142	5.8	2.7	5.1	1.9	Iris-virginica
143	6.8	3.2	5.9	2.3	Iris-virginica
144	6.7	3.3	5.7	2.5	Iris-virginica

sehr häufig im Bereich des Machine Learning (ML) und DL vorkommen.

Sample Ein Sample ist ein komplettes Beispiel aus einem Datensatz. In den meisten Fällen ist ein Sample mit einer Reihe in einem Datensatz gleichzusetzen. In dem Fall des IDS entspricht ein Sample einer Blume mit ihren Eigenschaften und ihrer Gattung.

Feature Ein Feature beschreibt die Eigenschaften eines Samples. Im Falle des IDS hat jedes Sample vier Features.

Label Das Label drückt die Art des Samples aus, eine Zugehörigkeit zu einer Klasse. Im Falle des IDS besitzt jedes Sample ein Label. Im ganzen Datensatz gibt es drei Labels, und zwar die drei Blumengattungen.

B. Python

Laut Tiobe Index Juni 2018 ist Python die viert häufigst verwendete Programmiersprache [vgl. 2]. Python gilt als sehr einfach zu lernen und zu schreibende Programmiersprache. Deshalb ist die Python in sehr vielen Disziplinen der Informationstechnik (IT) vertreten, besonders auch im Bereich des Scientific Computing. Das die meisten großen DL-Frameworks Python als erste Programmiersprache anbieten zeigt, dass Python sich auch hier durchsetzen konnte. Alle Code-Beispiele in dieser Arbeit sind in Python geschrieben.

¹<https://archive.ics.uci.edu/ml/datasets/iris>

C. Python Bibliotheken

Im Bereich des DL gibt es eine Reihe von Python Bibliotheken, deren Einsatz stark zu empfehlen ist:

numpy Numpy ist die Standard Python Bibliothek für Matrizen Operationen. Numpy ist hoch optimiert und extrem schnell.

pandas Pandas ist eine Python Bibliothek Dataframe Objekte zur Verfügung stellt, mit denen Daten leicht gelesen, bearbeitet und gespeichert werden können. Unter anderem können Comma Separated Values (CSV) Dateien gelesen werden.

sklearn Sklearn ist eine ML Python Bibliothek die viele nützliche Funktionen zu dem Vorbereiten und Verarbeiten der Datensätze mitbringt.

matplotlib TODO

D. Deep Learning Frameworks

TODO

E. Keras

TODO

IV. DAS NEURON

A. Das Perceptron

B. Das ADALINE

C. Aktivierungsfunktionen

D. Fehlerfunktionen

E. Gradient Descent

F. Stochastic und Batch Gradient Descent

V. KÜNSTLICHE NEURONALE NETZE

A. Definition

B. Arten des Deep Learning

C. Vom Neuron zum Deep Learning

D. Stochastic und Batch Gradient Descent

E. Backpropagation

F. Optimizer

G. Das Densenet

H. Das Convolutional Neural Network

I. Das Recurrent Neural Network

LITERATUR

- [1] *Iris flower data set*. URL: https://en.wikipedia.org/wiki/Iris_flower_data_set (besucht am 29.06.2018).
- [2] *TIOBE Index June 2018*. URL: <https://www.tiobe.com/tiobe-index/> (besucht am 29.06.2018).