

-Machine Learning: Boosting-Algorithmen-

Seminararbeit

Student:	David Erdös	67906
Universität:	Hochschule Karlsruhe – Technik und Wirtschaft	
Studiengang:	Informatik Bachelor	
Semester:	Wintersemester 2023	
Dozent:	Prof. Dr. Baier	
Bearbeitet am:	1. Dezember 2023	

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation und Zielsetzung	1
1.2	Struktur der Arbeit	1
1.3	Ziel und Umfang der Arbeit	1
2	Grundlagen des Machine Learning (3-4 Seiten)	2
2.1	Modern Approaches in Machine Learning	2
2.2	Role of Boosting Algorithms in ML	2
2.3	Boosting Algorithms in Tabular Data Analysis	2
3	Boosting	3
3.1	Was ist Boosting	3
4	AdaBoost (5-6 Seiten)	4
4.1	Theoretische Grundlagen	4
4.2	Algorithmus-Struktur und Funktionsweise	4
4.3	Beispielanwendung mit Erläuterung	4
5	Gradient Boosting (5-6 Seiten)	5
5.1	Theoretische Grundlagen	5
5.2	Algorithmus-Struktur und Funktionsweise	5
5.3	Beispielanwendung mit Erläuterung	5
6	Vergleich von AdaBoost und Gradient Boosting (4-5 Seiten)	6
6.1	Gemeinsamkeiten und Unterschiede	6
6.2	Performance-Analyse in Benchmarks	6
6.3	Anwendungsbeispiele und Fallstudien	6
7	Aktuelle Trends und Entwicklungen (2-3 Seiten)	7
7.1	Neueste Forschungsergebnisse	7
7.2	Zukünftige Potenziale von Boosting-Algorithmen	7
8	Fazit und Ausblick (2-3 Seiten)	8
8.1	Zusammenfassung der Erkenntnisse	8
8.2	Reflexion über die Bedeutung für die Praxis	8
8.3	Ausblick auf zukünftige Forschungsthemen	8
	Literaturverzeichnis	10
	Abbildungsverzeichnis	11
	Tabellenverzeichnis	12

1 Einleitung

In der dynamischen Welt des maschinellen Lernens (ML) haben sich Boosting-Algorithmen als revolutionär erwiesen, indem sie beeindruckende Leistungen in einer Vielzahl von Anwendungen erbringen [1, Kapitel 1.1.2]. Besonders bei tabellarischen Datensätzen erzielen sie in fast allen Bereichen die besten Ergebnisse.

In dieser Seminararbeit liegt der Schwerpunkt auf AdaBoost und GradientBoosting, die gemäß Géron [2, S. 192] zu den bekanntesten Vertretern der Boosting-Familie gehören. Die Arbeit wird die beiden Algorithmen detailliert untersuchen und verständlich darstellen.

1.1 Motivation und Zielsetzung

Die Motivation für diese Arbeit liegt in der zunehmenden Bedeutung, die Boosting-Algorithmen in der modernen Datenanalyse zukommt. Insbesondere in Situationen, in denen tabellenartige Datensätze verarbeitet werden, haben sich diese Algorithmen als besonders effektiv erwiesen. Dies spiegelt sich auch in ihrer Beliebtheit in Data-Science-Wettbewerben und realen Anwendungsbeispielen wider. Das Ziel dieser Seminararbeit ist es, ein tiefes Verständnis für die Funktionsweise von AdaBoost und GradientBoosting zu entwickeln und deren Einsatzmöglichkeiten anhand konkreter Beispiele zu demonstrieren. Hierbei wird besonderer Wert darauf gelegt, die zugrundeliegenden Konzepte in einer für Bachelorstudenten der Informatik im 6. Semester verständlichen Form darzulegen.

1.2 Struktur der Arbeit

Die Seminararbeit gliedert sich in mehrere Abschnitte, die schrittweise aufeinander aufbauen. Zunächst wird ein Überblick über die Grundlagen des maschinellen Lernens gegeben, gefolgt von einer detaillierten Betrachtung der Boosting-Algorithmen, insbesondere AdaBoost und GradientBoosting. Anschließend erfolgt ein Vergleich dieser beiden Methoden, wobei deren Stärken und Schwächen in verschiedenen Anwendungsszenarien beleuchtet werden. Abschließend wird ein Ausblick auf zukünftige Entwicklungen und mögliche Forschungsrichtungen im Bereich des Boostings gegeben.

[3, text] [1, text] [2, text] [4, text] [5, text]

1.3 Ziel und Umfang der Arbeit

2 Grundlagen des Machine Learning (3-4 Seiten)

2.1 Modern Approaches in Machine Learning

Überblick über aktuelle Trends und Innovationen im Machine Learning Vorstellung fortgeschrittener Techniken und Methoden Diskussion über die Bedeutung von Deep Learning und künstlichen neuronalen Netzen

2.2 Role of Boosting Algorithms in ML

Einführung in Boosting-Algorithmen und ihre Relevanz Spezifische Betrachtung von AdaBoost und Gradient Boosting Vergleich von Boosting-Algorithmen mit anderen fortgeschrittenen Methoden

2.3 Boosting Algorithms in Tabular Data Analysis

Bedeutung von tabellenartigen Datensätzen in fortgeschrittenen ML-Anwendungen Analyse, wie AdaBoost und Gradient Boosting bei tabellenartigen Daten effektiv sind Fallstudien und Beispiele aus der Praxis, die den Einsatz dieser Algorithmen zeigen

3 Boosting

Aus meiner eigenen Erfahrung weiß ich, dass es selten besonders positive Effekte hat, zusammen mit Kommilitonen eine Klausur zu lösen. Im Regelfall wird es einen Studenten geben, der sich relativ zur Gruppe, am besten auskennt und durch sein überdurchschnittliches Engagement, die Gruppenleistung auf etwa das Niveau bringt, was seiner Eigenleistung entsprochen hätte.

In einem extremeren Fall, bei dem alle Studenten sich wenig Wissen zum Thema angeeignet haben, dass ihre Antworten nur geringfügig besser sind als zufälliges Raten, gibt es die realistische Chance, dass die Gruppenleistung sogar schlechter ausfällt, als hätte man tatsächlich geraten.

Die Vorstellung, dass durch Zusammenarbeit von Personen mit wenig Wissen Ergebnisse erzielt werden, die weit über der Durchschnittsleistung, als auch der individuellen Bestleistung liegen, scheint für Menschen sehr unrealistisch. Sogar die Bibel warnt metaphorisch vor der Zusammenarbeit zweier für eine Aufgabenstellung nicht geeigneter Personen. Es heißt: ‘Wenn aber ein Blinder den andern führt, so fallen sie beide in die Grube.’ (Mt 23,16; Mt 23,24; Lk 6,39; Röm 2,19). Umso mehr überraschender scheint es, dass gerade im Bereich des Machine Learning die Zusammenarbeit schwacher Modelle zu erstaunlich starken Modellen kombiniert werden können.

3.1 Was ist Boosting

4 AdaBoost (5-6 Seiten)

4.1 Theoretische Grundlagen

4.2 Algorithmus-Struktur und Funktionsweise

4.3 Beispielanwendung mit Erläuterung

5 Gradient Boosting (5-6 Seiten)

5.1 Theoretische Grundlagen

5.2 Algorithmus-Struktur und Funktionsweise

5.3 Beispielanwendung mit Erläuterung

6 Vergleich von AdaBoost und Gradient Boosting (4-5 Seiten)

6.1 Gemeinsamkeiten und Unterschiede

6.2 Performance-Analyse in Benchmarks

6.3 Anwendungsbeispiele und Fallstudien

7 Aktuelle Trends und Entwicklungen (2-3 Seiten)

7.1 Neueste Forschungsergebnisse

7.2 Zukünftige Potenziale von Boosting-Algorithmen

8 Fazit und Ausblick (2-3 Seiten)

8.1 Zusammenfassung der Erkenntnisse

8.2 Reflexion über die Bedeutung für die Praxis

8.3 Ausblick auf zukünftige Forschungsthemen

Glossar

AdaBoost Ein Machine Learning-Algorithmus, der auf dem Prinzip des Boosting basiert.. 1, 2

Gradient Boosting Eine Methode des maschinellen Lernens, die für Regression und Klassifikation verwendet wird.. 1, 2

Literaturverzeichnis

- [1] R. E. Schapire und Y. Freund, *Boosting : foundations and algorithms* (Adaptive Computation and Machine Learning Series). Cambridge, Mass. [u.a.]: MIT Press, 2012, Includes bibliographical references and index, ISBN: 9780262017183. Adresse: <http://www.gbv.de/dms/ilmenau/toc/672276232.PDF%20;%20https://zbmath.org/?q=an:1278.68021>.
- [2] A. [Géron, *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow : Konzepte, Tools und Techniken für intelligente Systeme*, 1. Auflage, K. [Rother, Hrsg. Heidelberg: O'Reilly, 2018, Äuthorized German translation of the English edition of 'Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems', ISBN 978-1-491-96228-9, (c) 2017 Impressum; ÜS-Bestseller zu Deep Learning Umschlag"Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem Imprint Ö'Reilly". Rückseite der Titelseite. - "Deutschsprachige O'Reilly-Bücher werden vom dpunkt.verlag in Heidelberg publiziert, vermarktet und vertrieben.- Webseite www.oreilly.de; 201712; aa, ISBN: 9783960090618. Adresse: http://www.gbv.de/dms/ilmenau/toc/898831717.PDF%20;%20http://vub.de/cover/data/isbn%3A9783960090618/medium/true/de/vub/cover.jpg%20;%20http://deposit.d-nb.de/cgi-bin/dokserv?id=6ba05f7dfc5242d6bbc6f653e199e877&prov=M&dok_var=1&dok_ext=htm%20;%20https://www.oreilly.de/buecher/13111/9783960090618-praxiseinstieg-machine-learning-mit-scikit-learn-und-tensorflow.html.
- [3] J. [Frochte, *Maschinelles Lernen : Grundlagen und Algorithmen in Python* (Hanser eLibrary), 3., überarbeitete und erweiterte Auflage. München: Hanser, 2020, ISBN: 9783446463554. Adresse: <https://dx.doi.org/10.3139/9783446463554%20;%20https://doi.org/10.3139/9783446463554%20;%20https://www.hanser-elibrary.com/doi/book/10.3139/9783446463554%20;%20http://dx.doi.org/10.3139/9783446463554>.
- [4] G. [James, *An introduction to statistical learning : with applications in Python* (Springer texts in statistics), D. [Witten, T. [Hastie, R. [Tibshirani und J. E. [Taylor, Hrsg. Cham, Switzerland: Springer, 2023, ISBN: 9783031387463.
- [5] T. [Hastie, *The elements of statistical learning : data mining, inference, and prediction* (Springer series in statistics), Second edition, R. [Tibshirani und J. H. [Friedman, Hrsg. New York, NY: Springer, 2009, Hier auch später erschienene, unveränderte Nachdrucke; Literaturverzeichnis: Seite 699-727, ISBN: 9780387848570. Adresse: <http://www.gbv.de/dms/ilmenau/toc/572093853.PDF%20;%20https://zbmath.org/?q=an:1273.62005%20;%20https://swbplus.bsz-bw.de/bsz287727726kap.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726vor.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726inh.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726cov.jpg>.

Abbildungsverzeichnis

Tabellenverzeichnis