

Layerwise Relevance Propagation For Deep Neural Networks

Abschlussbericht

vorgelegt von **Gruppe 4**

Theo Conrads, Robin Khling, Marc Bremser

am 21. Juli 2020

am Mathematischen Institut der
Universitt zu Kln

Erstgutachter: Prof. Dr. Axel Klawonn



Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	v
1 Einleitung	vii
2 Einfhrgung Convolutional Neural Networks	ix
3 Deep Taylor Decomposition	xi
3.1 Taylor Decomposition	xi
3.2 Relevance Models	xi
3.2.1 Das Min-Max-Model	xi
3.2.2 Das Training-Free Model	xi
4 Layerwise Relevance Propagation	xiii
4.1 Herleitung der Basisregel	xiii
4.2 Erweiterungen der Basisregel	xiii
5 Implementierung	xv
5.1 Implementierung des Min-Max-Models	xv
5.2 Implementierung LRP	xv
6 Ergebnisse	xvii
6.1 Ergebnisse Min-Max-Model	xvii
6.2 Ergebnisse LRP inkl. Training-Free	xvii

7 Fazit

xix

Literaturverzeichnis

xxi

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Hier grundsätzliche Einführung

2 Einführung Convolutional Neural Networks

Kurze Einführung in den Aufbau und die Funktion von CNNs (vllt. 2 Seiten)

3 Deep Taylor Decomposition

3.1 Taylor Decomposition

Einführung in die einfache Taylor Decomposition

3.2 Relevance Models

3.2.1 Das Min-Max-Model

Theorie zum Min-Max-Model

3.2.2 Das Training-Free Model

Theorie zum Training-Free Ansatz

4 Layerwise Relevance Propagation

4.1 Herleitung der Basisregel

Herleitung nach PLOS-Paper. Bezug zu DTD herstellen

4.2 Erweiterungen der Basisregel

LRP- ϵ , LRP- $\alpha - \beta$, LRP- γ

5 Implementierung

5.1 Implementierung des Min-Max-Models

Erklärung der wichtigsten Codeabschnitte. Ergebnisse werden voraussichtlich erst im Kapitel Ergebnisse präsentiert.

5.2 Implementierung LRP

Erklärung der wichtigsten Code-Abschnitte

6 Ergebnisse

6.1 Ergebnisse Min-Max-Model

Prsentation der Ergebnisse. Bilder/Plots (Zeitaufwand etc.)

6.2 Ergebnisse LRP inkl. Training-Free

Prsentation der Ergebnisse. Bilder/Plots (Zeitaufwand etc.)

7 Fazit

Erklärung warum das ganze Thema großer Schrott ist :-)

Literaturverzeichnis

- [1] A. Toselli and O. Widlund. *Domain Decomposition Methods- Algorithms and Theory*, volume 34. Springer Series in Computational Mathematics, 2005.