

Neuronale Netze

Proseminar Data Mining

Lukas Krenz
Fakultät für Informatik
Technische Universität München
Email: lukas@krenz.land

Zusammenfassung—Das Paper ist eine kurze Einführung in Neuronale Netze, insbesondere in Multilayer Perceptrons inklusive Analyse ihrer praktischen Anwendung.

Schlüsselworte—Neuronale Netze, Multilayer Perceptrons, Backpropagation

INHALTSVERZEICHNIS

| | | |
|------------------|---|---|
| I | Einleitung | 1 |
| II | Vom linearen Modell zum Neuronalen Netz | 1 |
| II-A | Das Perceptron als additatives lineares Modell | 1 |
| II-B | Das Perceptron als einfaches ANN . . | 1 |
| II-C | Das XOR-Problem: nicht linear separierbare Probleme | 1 |
| II-D | Die Lösung: Neuronale Netze | 1 |
| III | Neuronale Netze | 1 |
| III-A | Die Aktivierungsfunktion | 1 |
| III-B | Die Kostenfunktion | 2 |
| III-C | Die Minimierung der Kostenfunktion - Backpropagation | 2 |
| III-D | Die numerische Methode - Gradient Descent und verwandte | 2 |
| IV | Neuronale Netze in der Praxis | 2 |
| IV-A | Problematik von lokalen Minima . . . | 2 |
| IV-B | Anzahl Hidden Layer | 2 |
| IV-C | Initialisierung der Gewichte | 2 |
| IV-D | Vermeidung von Overfitting | 2 |
| IV-E | Regression | 2 |
| IV-F | Klassifizierung | 2 |
| IV-G | | 2 |
| V | Zusammenfassung und Ausblick | 2 |
| Literatur | | 2 |

I. EINLEITUNG

Neuronale Netze (ANN) ist ein graphisches, mathematisches Modell, dass für vielfältige Anwendungen:

- Regression
- Clustering

- Annäherung von beliebigen Funktionen eignet.

Auch, wenn sie auf den ersten Eindruck kompliziert und unverständlich wirken, ist die mathematische und statistische Grundlage simpel.

Im folgenden wird der Weg vom linearen Modell zum Perceptron zum ANN gezeigt - mit jedem Schritt wird dabei die Komplexität, aber auch die Flexibilität erweitert. Nur die sogenannten Feed-Forward NNs werden hier beschrieben, wie sich später zeigen wird, sind diese auch mächtig genug, um alle genannten Probleme zu lösen - auch wenn andere Methoden effizienter sein können.

II. VOM LINEAREN MODELL ZUM NEURONALEN NETZ

Sehr simpel; lineares Modell als Input -> Output layer, dann lineare Funktion als Hidden Layer -> Lineares Modell (Vergleich; Bild Logistic Regression vs Lineares Perceptron)

A. Das Perceptron als additatives lineares Modell

Modell perceptron.
Problem: OR geht, XOR nicht -> lineare separierbarkeit

B. Das Perceptron als einfaches ANN

Verallgemeinerung zum ANN.
blabla mit drei Quellenangaben [1]–[3]

C. Das XOR-Problem: nicht linear separierbare Probleme

Theorem: Single Layer kann alle Funktionen annähern.

D. Die Lösung: Neuronale Netze

Neuronale Netze sind eine verallgemeinerung von Perceptrons. Sie besitzen noch eine weitere Schicht Neuronen, die jedoch nicht nur lineare Transformationen durchführen, sondern zusätzlich noch eine nicht lineare.

Dadurch ist ein Fit nicht mehr analytisch einfach lösbar; das Training muss numerisch erfolgen.

III. NEURONALE NETZE

A. Die Aktivierungsfunktion

Nicht linear.
Sigmoidale Form.
Besonders effizient: Nicht symmetrische Form.

Sigmoid Function; Numerical Optim., ...

B. Die Kostenfunktion

Mean squared Errors(MSE) (euklidische Norm)

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (1)$$

oder negative Cross entropy (TODO: Quelle, richtige Formel)

$$-\frac{1}{N} \sum_{n=1}^N \left[y_n \log \hat{y}_n + (1 - y_n) \log(1 - \hat{y}_n) \right] \quad (2)$$

C. Die Minimierung der Kostenfunktion - Backpropagation

D. Die numerische Methode - Gradient Descent und verwandte

IV. NEURONALE NETZE IN DER PRAXIS

Weight Decay, Anfangswert

A. Problematik von lokalen Minima

Cost function hat mehrere Minima.

Lösung: Neu starten

Oft nicht globales minimum erwünscht (siehe Overfitting)

B. Anzahl Hidden Layer

C. Initialisierung der Gewichte

Randomisiert - oft nicht gut

D. Vermeidung von Overfitting

Weight Decay (entspricht L2-Regularisierung)

Bayesian View of L2-Reg.

E. Regression

F. Klassifizierung

Cross entropy.

Soft-Max-Funktion

G. ...

[https://colah.github.io/posts/2014-](https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/)

03-NN-Manifolds-Topology/

[https://snippyhollow.github.io/blog/2014/08/09/so-](https://snippyhollow.github.io/blog/2014/08/09/so-you-wanna-try-deep-learning/)

you-wanna-try-deep-learning/

<http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>

Efficient Backprop Zitat: Nonlinear activation functions are what give neural networks their nonlinear capabilities.

<https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw>

<https://stats.stackexchange.com/questions/63152/what-does-the-hidden-layer-in-a-neural-network-compute>

V. ZUSAMMENFASSUNG UND AUSBLICK

blabla

LITERATUR

- [1] B. Claise, "IPFIX protocol specifications," Internet-Draft, draft-ietf-ipfix-protocol-07, December 2004.
- [2] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, S. T. Kent, and W. T. Strayer, "Hash-based IP traceback," in *ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, 2001.
- [3] A. Belenky and N. Ansari, "IP traceback with deterministic packet marking," *IEEE Communications Letters*, vol. 7, no. 4, pp. 162–164, 2003.