



Universiteit  
Leiden  
The Netherlands

## Risk bounds for deep learning

Bos, J.M.

### Citation

Bos, J. M. (2024, June 19). *Risk bounds for deep learning*. Retrieved from <https://hdl.handle.net/1887/3763887>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3763887>

**Note:** To cite this publication please use the final published version (if applicable).

# Risk bounds for deep learning

Thijs Bos

**Risk bounds for deep learning**

**Author:** Thijs Bos

176 pages

PhD thesis, Universiteit Leiden, The Netherlands (2024)

With summary in English and Dutch

**ISBN: 978-94-6469-978-4**

Printed by ProefschriftMaken || [www.proefschriftmaken.nl](http://www.proefschriftmaken.nl)

The research described in this thesis was financially supported by the NWO/STAR grant 613.009.034b and the NWO Vidi grant VI.Vidi.192.021.

# Risk bounds for deep learning

## Proefschrift

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof.dr.ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op woensdag 19 juni 2024  
klokke 10:00 uur

door

Johan Matthijs Bos  
geboren te Bergschenhoek  
in 1991

**Promotores:**

Prof.dr. P.D. Grünwald (Universiteit Leiden en Centrum Wiskunde & Informatica)

Prof.dr. A.J. Schmidt-Hieber (Universiteit Twente)

**Promotiecommissie:**

Prof.dr.ir. G.L.A. Derks

Prof.dr. J.J. Goeman

Prof.dr. A. Rohde (Universität Freiburg)

Prof.dr. M. Kohler (Technische Universität Darmstadt)

Dr. R.M. Castro (Technische Universiteit Eindhoven)

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Statistical background . . . . .	1
1.1.1	Nonparametric statistics . . . . .	2
1.1.2	Loss, risk and minimax rates . . . . .	5
1.1.3	Empirical risk minimization . . . . .	7
1.1.4	Curse of dimensionality . . . . .	9
1.2	Deep learning . . . . .	10
1.2.1	Training of neural networks . . . . .	13
1.3	Introduction for Chapter 2: Classification . . . . .	14
1.4	Introduction for Chapter 3: Multivariate density estimation . . . . .	16
1.5	Introduction for Chapter 4: Optimization motivated by biological neural networks . . . . .	17
<b>2</b>	<b>Convergence rates of deep ReLU networks for multiclass classification</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	The multiclass classification model . . . . .	21
2.2.1	Deep ReLU networks . . . . .	24
2.3	Main Results . . . . .	25
2.3.1	Relationship with Hellinger distance . . . . .	31
2.3.2	Oracle Inequality . . . . .	33
2.4	Proofs . . . . .	35
2.4.1	Approximation related results . . . . .	36
2.4.2	Oracle inequality related results . . . . .	41
2.A	Basic network properties and operations . . . . .	51
2.A.1	Embedding properties of neural network function classes . . . . .	51
2.A.2	Scaling numbers . . . . .	52
2.A.3	Negative numbers . . . . .	53
2.B	Neural networks approximating the logarithm . . . . .	53

2.B.1	Taylor approximation . . . . .	54
2.B.2	Partition of unity . . . . .	57
2.C	Further technicalities . . . . .	64
<b>3</b>	<b>A supervised deep learning method for nonparametric density estimation</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.1.1	Notation . . . . .	75
3.2	Conversion into a supervised learning problem . . . . .	75
3.3	Main results . . . . .	77
3.3.1	Neural networks . . . . .	79
3.3.2	Structural constraints: compositions of functions . . . . .	79
3.4	Examples of multivariate densities with compositional structure . . . . .	82
3.4.1	Copulas . . . . .	84
3.4.2	Mixture distributions . . . . .	88
3.5	Simulations . . . . .	89
3.5.1	Methods . . . . .	89
3.5.2	Densities . . . . .	90
3.5.3	Neural network training setup . . . . .	93
3.5.4	Simulation Results . . . . .	93
3.6	Proofs for Section 3.3 . . . . .	97
3.6.1	Proof of Theorem 3.3.1 . . . . .	97
3.6.2	Proof of Theorem 3.3.4 . . . . .	101
3.7	Proofs for Section 3.4 . . . . .	102
3.7.1	Proof of Theorem 3.4.5 . . . . .	104
3.8	Proofs for Section 3.5 . . . . .	106
3.9	Proofs for Section 3.6 . . . . .	107
<b>4</b>	<b>Convergence guarantees for forward gradient descent in the linear regression model</b>	<b>125</b>
4.1	Introduction . . . . .	125
4.2	Weight-perturbed forward gradient descent . . . . .	127
4.3	Convergence rates in the linear regression model . . . . .	129
4.4	Proofs . . . . .	133
4.4.1	Proof of Theorem 4.3.2 . . . . .	134
4.4.2	Proof of Theorem 4.3.3 . . . . .	137
<b>5</b>	<b>General discussion</b>	<b>141</b>
5.1	Statistical theory and training of neural networks . . . . .	142
5.2	Model assumptions . . . . .	142

<i>Contents</i>	III
<b>Bibliography</b>	<b>158</b>
<b>Summary</b>	<b>159</b>
<b>Samenvatting</b>	<b>161</b>
<b>Acknowledgements</b>	<b>163</b>
<b>Curriculum Vitae</b>	<b>165</b>
<b>List of publications</b>	<b>167</b>