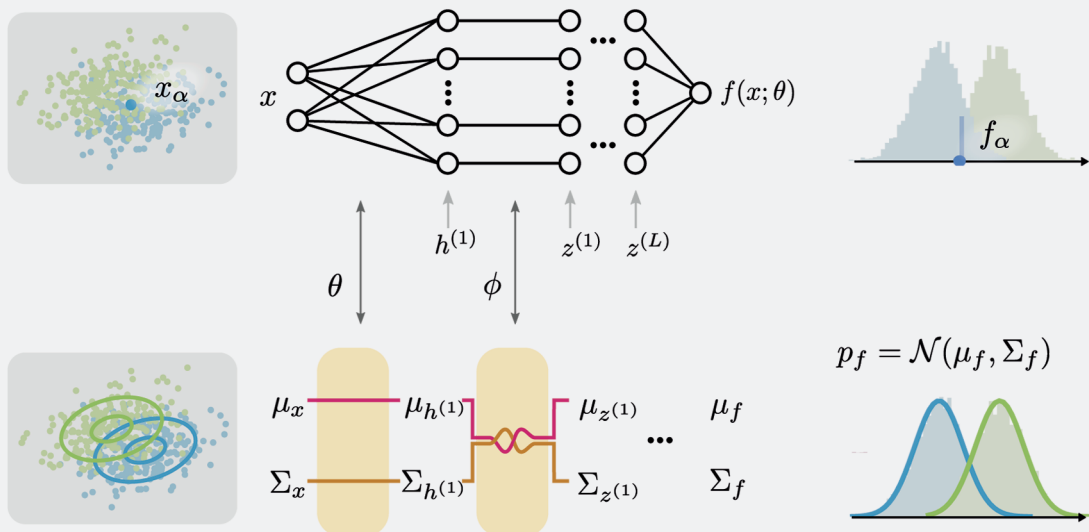




Mechanics of deep neural networks beyond the Gaussian limit

Kirsten Fischer

Information

Band / Volume 110

ISBN 978-3-95806-815-5

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)
Computational and Systems Neuroscience (IAS-6)

Mechanics of deep neural networks beyond the Gaussian limit

Kirsten Fischer

Schriften des Forschungszentrums Jülich
Reihe Information / Information

Band / Volume 110

ISSN 1866-1777

ISBN 978-3-95806-815-5

Contents

1	Introduction	1
2	Decomposing neural networks as mappings of correlation functions	5
2.1	Introduction	6
2.2	Setup	6
2.3	Theoretical Background	7
2.3.1	Empirical risk minimization	7
2.3.2	Parameterizing probability distributions by cumulants	8
2.4	Theory	11
2.4.1	Cumulants drive network training	11
2.4.2	Transformation of cumulants by the network	12
2.4.3	Input layer extracts higher-order cumulants from data	14
2.4.4	Statistical representation of feed-forward networks	16
2.5	Experiments	17
2.5.1	Training details	18
2.5.2	Statistical information encoding for the XOR task	18
2.5.3	Gaussian statistics are essential to the MNIST data set	23
2.5.4	Extracting higher-order cumulants in the input layer	25
2.5.5	High dimensionality of input data justifies Gaussian description of fully-connected deep networks	27
2.6	Conclusion	28
2.6.1	Limitations	29
2.6.2	Relation to other works	30

3	Critical feature learning in deep neural networks	33
3.1	Introduction	34
3.2	Setup	34
3.3	Theoretical background	35
3.3.1	Bayesian supervised learning	35
3.3.2	Neural Network Gaussian Process	37
3.3.3	Network prior in a field-theoretic formulation	37
3.3.4	Next-to-leading-order corrections	41
3.3.5	Criticality in neural networks	43
3.3.6	Large deviation theory	44
3.4	Theory	45
3.4.1	Large deviation approach for the posterior kernels	45
3.4.2	Forward-backward kernel propagation equations	48
3.4.3	Perturbative, leading-order solution of the forward-backward kernel equations	49
3.4.4	Fluctuation corrections lead to feature learning	49
3.5	Experiments	52
3.5.1	Comparative analysis with trained networks	53
3.5.2	Link between feature learning corrections and criticality	55
3.5.3	Downscaling network outputs boosts feature learning	57
3.6	Conclusion	58
3.6.1	Limitations	58
3.6.2	Relation to other works	59

4	Field theory for optimal signal propagation in residual networks	61
4.1	Introduction	62
4.2	Setup	63
4.3	Theoretical background	63
4.3.1	NNGP for ResNets	63
4.4	Theory	64
4.4.1	Network prior in a field-theoretic framework	65
4.4.2	Next-to-leading-order correction	72
4.5	Experiments	75
4.5.1	Kernels and response function in networks at initialization	75
4.5.2	Optimal scaling in residual networks	76
4.5.3	Optimal scaling depends strongly on depth hyperparameter	78
4.5.4	Optimal scaling across data sets	79
4.6	Conclusion	80
4.6.1	Limitations	80
4.6.2	Relation to other works	81
5	Discussion	83

Appendix	89
A Decomposing neural networks as mappings of correlation functions	89
A.1 Interaction functions for different activation functions	89
A.1.1 ReLU	89
A.1.2 Quadratic activation function	92
A.2 Higher-order cumulants of post-activations from weak correlations . .	93
A.3 Suppression of higher-order cumulants in wide networks	96
A.4 Networks with quadratic activation function	100
A.5 Depth scales of signal propagation in neural networks	100

B	Critical feature learning in deep neural networks	109
B.1	Conjugate kernels yield training error	109
B.2	Deep linear networks	110
B.3	Target-kernel-adaptation in linear networks	113
B.4	Relation to the Neural Tangent Kernel	115
B.5	Langevin stochastic gradient descent	118
B.6	Additional details on numerical evaluation of theory	120
B.6.1	Weight variance of the input layer	120
B.6.2	Gaussian integrals	120
B.6.3	Annealing in network width	120
C	Field theory for optimal signal propagation in residual networks	123
C.1	Maximum entropy condition for optimal scaling	123
C.2	Supplementary figures	124
C.2.1	Sub-exponential decay of response function	124
C.2.2	Normalized input kernels	124

Bibliography	127
---------------------	------------

Information
Band / Volume 110
ISBN 978-3-95806-815-5