

Contents

Introduction — V

Acknowledgement — XI

1	Objective functions, neural networks, and linear algebra — 1
1.1	Notations — 1
1.2	The least squares method — 2
1.2.1	Recursivity — 6
1.2.2	Nonlinearity — 6
1.2.3	Functions and Neural Networks — 7
1.3	Logistic function, cross-entropy, classification, CNNs — 12
1.3.1	Logistic function, cross-entropy — 12
1.3.2	Classification — 15
1.3.3	Convolutional neural networks — 19
1.4	Summary of the chapter — 26
2	Approximation properties — 27
2.1	Functional spaces — 27
2.2	Cybenko theorem — 28
2.3	Simple constructive proofs — 30
2.3.1	Approximation in dimension one — 30
2.3.2	Approximation in higher dimensions — 35
2.4	Connection with finite element theory — 38
2.4.1	1D piecewise affine finite elements — 38
2.4.2	2D piecewise affine finite elements — 41
2.5	ReLU function, power of depth — 45
2.5.1	An interesting series — 45
2.5.2	Yarotsky theorem — 52
2.5.3	Takagi function — 58
2.6	Summary of the chapter — 64
3	A functional equation — 65
3.1	A contractive functional equation — 65
3.1.1	Two examples — 66
3.1.2	Number of intervals — 67
3.1.3	Nonpolynomial solutions — 67
3.1.4	Construction of polynomial solutions — 67
3.1.5	Fixed point algorithm — 70

3.2	Application to Neural Networks —	71
3.2.1	A first Neural Network implementation —	71
3.2.2	A second Neural Network implementation —	74
3.2.3	Time interpretation of Neural Network implementations —	77
3.3	Summary of the chapter —	79
4	Datasets —	81
4.1	The curse of dimension and uniform sampling —	82
4.2	The Monte Carlo method —	85
4.3	The Latin hypercube method —	90
4.4	Comparison of uniform sampling versus Latin hypercube method —	96
4.5	Lattice rules —	101
4.6	Numerical application —	104
4.7	Summary of the chapter —	106
5	Stochastic gradient methods —	107
5.1	Steepest gradient method —	108
5.2	Momentum methods and ADAM —	113
5.3	Batches —	118
5.3.1	A stochastic steepest gradient method —	118
5.3.2	Convergence —	119
5.4	Initialization —	122
5.5	A theoretical control of overfitting —	123
5.6	The regularity problem —	124
5.7	Summary of the chapter —	130
6	Examples and research in the field —	131
6.1	A CNN for the MNIST database —	131
6.2	A CNN for the CIFAR database —	131
6.3	Classification of numerical data produced by a PDE solver —	132
6.4	Oscillating curves for a classification session on the computer —	134
6.5	Training of simple polynomials —	135
6.6	Construction of simple PDE solvers —	136
6.7	Finite Volume numerical fluxes for interface reconstruction in fluids —	139
6.8	Autoencoders —	142
6.9	Research in the field —	145
6.9.1	More on NNs and scientific computing —	145
6.9.2	PINNs —	146
6.9.3	Tensor trains —	147
6.9.4	Learning and optimal transport —	147

Bibliography — 149

Index — 155