## **Contents**

1	About this book —— 1
2	Introduction to machine learning: what and why? —— 2
2.1	Some motivation —— 2
2.2	What is machine learning? —— 2
3	Classification problem —— 4
4	The fundamentals of artificial neural networks —— 6
4.1	Basic definitions —— 6
4.2	ANN classifiers and the softmax function —— 10
4.3	The universal approximation theorem —— 12
4.4	Why is non-linearity in ANNs necessary? —— 14
4.4.1	0 + 0 = 8? —— 14
4.4.2	Non-linear activation functions are necessary in ANNs —— 15
4.5	Exercises —— 16
5	Supervised, unsupervised, and semisupervised learning —— 19
5.1	Basic definitions —— 19
5.2	Example of unsupervised learning: detecting bank fraud —— 21
5.3	Exercises —— 22
6	The regression problem —— 24
6.1	What is regression? How does it relate to ANNs? —— 24
6.2	Example: linear regression in dimension 1 —— 26
6.3	Logistic regression as a single neuron ANN —— 27
6.3.1	1D example: studying for an exam —— 27
6.3.2	2D example of admittance to graduate school: separation of sets and decision boundary —— <b>30</b>
6.3.3	Relation between ANNs and regression —— 32
6.3.4	Logistic regression vs. networks with many layers —— 33
6.4	Exercises —— 35
7	Support vector machine —— 40
7.1	Preliminaries: convex sets and their separation, geometric Hahn-Banach
	theorem —— 40
7.2	Support vector machine —— 42
7.3	Hard margin SVM classifiers —— 44
7.4	Soft margin SVM classifier —— 46
7.5	Exercises —— 49



8	Gradient descent method in the training of DNNs —— 52
8.1	Deterministic gradient descent for the minimization of multivariable
	functions — 53
8.2	Additive loss functions — 56
8.3	What are SGD algorithms? When to use them? —— 57
8.4	Epochs in SGD —— <b>58</b>
8.5	Weights —— <b>59</b>
8.6	Choosing the batch size through a numerical example —— 60
8.7	Exercises —— 64
9	Backpropagation —— 67
9.1	Computational complexity —— <b>67</b>
9.2	Chain rule review —— 72
9.3	Diagrammatic representation of the chain rule in simple examples —— 74
9.4	The case of a simple DNN with one neuron per layer —— 78
9.5	Backpropagation algorithm for general DNNs —— 83
9.6	Exercises —— 88
10	Convolutional neural networks —— 93
10.1	Convolution —— 93
10.1.1	Convolution of functions —— 93
10.1.2	Convolution of matrices —— 94
10.1.3	Hadamard product and feature detection —— 96
10.2	Convolutional layers —— 99
10.3	Padding layer —— <b>102</b>
10.4	Pooling layer —— 103
10.5	Building CNNs —— 105
10.6	Equivariance and invariance —— 107
10.7	Summary of CNNs —— 114
10.8	Exercises —— 115

A Review of the chain rule —— 119

Bibliography —— 121

Index —— 125