

Contents

<i>Foreword</i>	ix
<i>Preface</i>	xi
1 The framework of learning	1
1.1 Introduction	1
1.2 A formal setting	5
1.3 Hypothesis spaces and target functions	9
1.4 Sample, approximation, and generalization errors	11
1.5 The bias–variance problem	13
1.6 The remainder of this book	14
1.7 References and additional remarks	15
2 Basic hypothesis spaces	17
2.1 First examples of hypothesis space	17
2.2 Reminders I	18
2.3 Hypothesis spaces associated with Sobolev spaces	21
2.4 Reproducing Kernel Hilbert Spaces	22
2.5 Some Mercer kernels	24
2.6 Hypothesis spaces associated with an RKHS	31
2.7 Reminders II	33
2.8 On the computation of empirical target functions	34
2.9 References and additional remarks	35
3 Estimating the sample error	37
3.1 Exponential inequalities in probability	37
3.2 Uniform estimates on the defect	43
3.3 Estimating the sample error	44
3.4 Convex hypothesis spaces	46
3.5 References and additional remarks	49

4	Polynomial decay of the approximation error	54
4.1	Reminders III	55
4.2	Operators defined by a kernel	56
4.3	Mercer's theorem	59
4.4	RKHSs revisited	61
4.5	Characterizing the approximation error in RKHSs	63
4.6	An example	68
4.7	References and additional remarks	69
5	Estimating covering numbers	72
5.1	Reminders IV	73
5.2	Covering numbers for Sobolev smooth kernels	76
5.3	Covering numbers for analytic kernels	83
5.4	Lower bounds for covering numbers	101
5.5	On the smoothness of box spline kernels	106
5.6	References and additional remarks	108
6	Logarithmic decay of the approximation error	109
6.1	Polynomial decay of the approximation error for \mathcal{C}^∞ kernels	110
6.2	Measuring the regularity of the kernel	112
6.3	Estimating the approximation error in RKHSs	117
6.4	Proof of Theorem 6.1	125
6.5	References and additional remarks	125
7	On the bias–variance problem	127
7.1	A useful lemma	128
7.2	Proof of Theorem 7.1	129
7.3	A concrete example of bias–variance	132
7.4	References and additional remarks	133
8	Least squares regularization	134
8.1	Bounds for the regularized error	135
8.2	On the existence of target functions	139
8.3	A first estimate for the excess generalization error	140
8.4	Proof of Theorem 8.1	148
8.5	Reminders V	151
8.6	Compactness and regularization	151
8.7	References and additional remarks	155
9	Support vector machines for classification	157
9.1	Binary classifiers	159

9.2	Regularized classifiers	161
9.3	Optimal hyperplanes: the separable case	166
9.4	Support vector machines	169
9.5	Optimal hyperplanes: the nonseparable case	171
9.6	Error analysis for separable measures	173
9.7	Weakly separable measures	182
9.8	References and additional remarks	185
10	General regularized classifiers	187
10.1	Bounding the misclassification error in terms of the generalization error	189
10.2	Projection and error decomposition	194
10.3	Bounds for the regularized error $\mathcal{D}(\gamma, \phi)$ of f_γ	196
10.4	Bounds for the sample error term involving f_γ	198
10.5	Bounds for the sample error term involving $f_{z, \gamma}^\phi$	201
10.6	Stronger error bounds	204
10.7	Improving learning rates by imposing noise conditions	210
10.8	References and additional remarks	211
	<i>References</i>	214
	<i>Index</i>	222