

# Contents

Preface	v
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Multilayer perceptron neural networks . . . . .	1
1.2 Regression and classification . . . . .	4
1.3 Learning and generalization . . . . .	7
1.3.1 Weight decay and effective number of parameters . . . . .	7
1.3.2 Ridge regression . . . . .	9
1.3.3 Bayesian learning . . . . .	11
1.4 Principles of pattern recognition . . . . .	14
1.4.1 Bayes rule and optimal classifier under Gaussian assumptions . . . . .	14
1.4.2 Receiver operating characteristic . . . . .	19
1.5 Dimensionality reduction methods . . . . .	20
1.6 Parametric versus non-parametric approaches and RBF networks	22
1.7 Feedforward versus recurrent network models . . . . .	24
<b>Chapter 2 Support Vector Machines</b>	<b>29</b>
2.1 Maximal margin classification and linear SVMs . . . . .	29
2.1.1 Margin . . . . .	29
2.1.2 Linear SVM classifier: separable case . . . . .	30
2.1.3 Linear SVM classifier: non-separable case . . . . .	34
2.2 Kernel trick and Mercer condition . . . . .	36
2.3 Nonlinear SVM classifiers . . . . .	37
2.4 VC theory and structural risk minimization . . . . .	44

2.4.1	Empirical risk versus generalization error . . . . .	44
2.4.2	Structural risk minimization . . . . .	48
2.5	SVMs for function estimation . . . . .	51
2.5.1	SVM for linear function estimation . . . . .	51
2.5.2	SVM for nonlinear function estimation . . . . .	54
2.5.3	VC bound on generalization error . . . . .	56
2.6	Modifications and extensions . . . . .	60
2.6.1	Kernels . . . . .	60
2.6.2	Extension to other convex cost functions . . . . .	63
2.6.3	Algorithms . . . . .	65
2.6.4	Parametric versus non-parametric approaches . . . . .	68

### **Chapter 3 Basic Methods of Least Squares Support Vector Machines** **71**

3.1	Least Squares Support Vector Machines for classification . . . . .	71
3.2	Multi-class formulations . . . . .	76
3.3	Link with Fisher discriminant analysis in feature space . . . . .	81
3.3.1	Linear Fisher discriminant analysis . . . . .	82
3.3.2	Fisher discriminant analysis in feature space . . . . .	85
3.4	Solving the LS-SVM KKT system . . . . .	86
3.4.1	Conjugate gradient iterative methods . . . . .	86
3.4.2	LS-SVM classifiers: UCI benchmarking results . . . . .	89
3.5	Least Squares Support Vector Machines for function estimation . . . . .	98
3.6	Links with regularization networks and Gaussian processes . . . . .	100
3.6.1	RKHS and reproducing property . . . . .	101
3.6.2	Representer theorem and regularization networks . . . . .	104
3.6.3	Gaussian processes . . . . .	106
3.7	Sparseness by pruning . . . . .	111

### **Chapter 4 Bayesian Inference for LS-SVM Models** **117**

4.1	Bayesian inference for LS-SVM classifiers . . . . .	117
4.1.1	Definition of network model, parameters and hyperparameters . . . . .	117
4.1.2	Bayes rule and levels of inference . . . . .	119
4.1.3	Probabilistic interpretation of LS-SVM classifiers (Level 1) . . . . .	122
4.1.4	Inference of the hyperparameters (Level 2) . . . . .	129

4.1.5	Inference of kernel parameters and model comparison (Level 3) . . . . .	131
4.1.6	Design of the LS-SVM classifier in the Bayesian evidence framework . . . . .	134
4.1.7	Example and benchmarking results . . . . .	136
4.2	Bayesian inference for LS-SVM regression . . . . .	139
4.2.1	Probabilistic interpretation of LS-SVM regressors (Level 1): predictive mean and error bars . . . . .	140
4.2.2	Inference of hyperparameters (Level 2) . . . . .	143
4.2.3	Inference of kernel parameters and model comparison (Level 3) . . . . .	144
4.3	Input selection by automatic relevance determination . . . . .	144
<b>Chapter 5 Robustness</b>		<b>149</b>
5.1	Noise model assumptions and robust statistics . . . . .	149
5.2	Weighted LS-SVMs . . . . .	150
5.3	Robust cross-validation . . . . .	156
5.3.1	$M$ -Estimators . . . . .	158
5.3.2	$L$ -Estimators . . . . .	163
5.3.3	Efficiency-robustness trade-off . . . . .	165
5.3.4	A robust and efficient cross-validation score function . . . . .	167
<b>Chapter 6 Large Scale Problems</b>		<b>173</b>
6.1	Low rank approximation methods . . . . .	173
6.1.1	Nyström method . . . . .	173
6.1.2	Incomplete Cholesky factorization . . . . .	176
6.2	Fixed Size LS-SVMs . . . . .	178
6.2.1	Estimation in primal weight space . . . . .	179
6.2.2	Active selection of support vectors . . . . .	181
6.3	Basis construction in the feature space . . . . .	183
6.4	Combining submodels . . . . .	190
6.4.1	Committee network approach . . . . .	190
6.4.2	Multilayer networks of LS-SVMs . . . . .	196
<b>Chapter 7 LS-SVM for Unsupervised Learning</b>		<b>201</b>
7.1	Support Vector Machines and linear PCA analysis . . . . .	201
7.1.1	Classical principal component analysis formulation . . . . .	202
7.1.2	Support vector machine formulation to linear PCA . . . . .	202

7.1.3	Including a bias term . . . . .	205
7.1.4	The reconstruction problem . . . . .	208
7.2	An LS-SVM approach to kernel PCA . . . . .	210
7.3	Links with density estimation . . . . .	214
7.4	Kernel CCA . . . . .	219
7.4.1	Classical canonical correlation analysis formulation . . . . .	219
7.4.2	Support vector machine formulation to linear CCA . . . . .	220
7.4.3	Extension to kernel CCA . . . . .	222
<b>Chapter 8 LS-SVM for Recurrent Networks and Control</b>		<b>225</b>
8.1	Recurrent Least Squares Support Vector Machines . . . . .	225
8.1.1	From feedforward to recurrent LS-SVM formulations . . . . .	225
8.1.2	A simplification to the problem . . . . .	228
8.1.3	Example: trajectory learning of chaotic systems . . . . .	230
8.2	LS-SVMs and optimal control . . . . .	232
8.2.1	The $N$ -stage optimal control problem . . . . .	232
8.2.2	LS-SVMs as controllers within the $N$ -stage optimal control problem . . . . .	234
8.2.3	Alternative formulation and stability issues . . . . .	238
8.2.4	Illustrative examples . . . . .	239
Appendix A		249
Bibliography		269
List of Symbols . . . . .		287
Acronyms . . . . .		289
Index		291