

# Support Vector Machines

Optimization Based Theory,  
Algorithms, and Extensions

Naiyang Deng  
Yingjie Tian  
Chunhua Zhang



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an Informa business

A CHAPMAN & HALL BOOK

---

# Contents

List of Figures	xxvii
List of Tables	xxi
Preface	xxiii
List of Symbols	xxvii
<b>1 Optimization</b>	<b>1</b>
1.1 Optimization Problems in Euclidian Space . . . . .	1
1.1.1 An example of optimization problems . . . . .	1
1.1.2 Optimization problems and their solutions . . . . .	3
1.1.3 Geometric interpretation of optimization problems . . . . .	4
1.2 Convex Programming in Euclidean Space . . . . .	5
1.2.1 Convex sets and convex functions . . . . .	6
1.2.1.1 Convex sets . . . . .	6
1.2.1.2 Convex functions . . . . .	7
1.2.2 Convex programming and their properties . . . . .	8
1.2.2.1 Convex programming problems . . . . .	8
1.2.2.2 Basic properties . . . . .	9
1.2.3 Duality theory . . . . .	12
1.2.3.1 Derivation of the dual problem . . . . .	12
1.2.3.2 Duality theory . . . . .	13
1.2.4 Optimality conditions . . . . .	15
1.2.5 Linear programming . . . . .	16
1.3 Convex Programming in Hilbert Space . . . . .	18
1.3.1 Convex sets and Fréchet derivative . . . . .	18
1.3.2 Convex programming problems . . . . .	19
1.3.3 Duality theory . . . . .	20
1.3.4 Optimality conditions . . . . .	20
*1.4 Convex Programming with Generalized Inequality Constraints in Euclidian Space . . . . .	21
1.4.1 Convex programming with generalized inequality con- straints . . . . .	21
1.4.1.1 Cones . . . . .	21
1.4.1.2 Generalized inequalities . . . . .	21

1.4.1.3	Convex programming with generalized inequality constraints . . . . .	22
1.4.2	Duality theory . . . . .	23
1.4.2.1	Dual cones . . . . .	23
1.4.2.2	Derivation of the dual problem . . . . .	23
1.4.2.3	Duality theory . . . . .	25
1.4.3	Optimality conditions . . . . .	26
1.4.4	Second-order cone programming . . . . .	27
1.4.4.1	Second-order cone programming and its dual problem . . . . .	27
1.4.4.2	Software for second-order cone programming . . . . .	30
1.4.5	Semidefinite programming . . . . .	31
1.4.5.1	Semidefinite programming and its dual problem . . . . .	31
1.4.5.2	Software for semidefinite programming . . . . .	35
*1.5	Convex Programming with Generalized Inequality Constraints in Hilbert Space . . . . .	36
1.5.1	$K$ -convex function and Fréchet derivative . . . . .	36
1.5.2	Convex programming . . . . .	36
1.5.3	Duality theory . . . . .	37
1.5.4	Optimality conditions . . . . .	38
<b>2</b>	<b>Linear Classification</b> . . . . .	<b>41</b>
2.1	Presentation of Classification Problems . . . . .	41
2.1.1	A sample (diagnosis of heart disease) . . . . .	41
2.1.2	Classification problems and classification machines . . . . .	43
2.2	Support Vector Classification (SVC) for Linearly Separable Problems . . . . .	45
2.2.1	Maximal margin method . . . . .	45
2.2.1.1	Derivation of the maximal margin method . . . . .	45
2.2.1.2	Properties of the maximal margin method . . . . .	48
2.2.2	Linearly separable support vector classification . . . . .	50
2.2.2.1	Relationship between the primal and dual problems . . . . .	50
2.2.2.2	Linearly separable support vector classification . . . . .	53
2.2.3	Support vector . . . . .	54
2.3	Linear $C$ -Support Vector Classification . . . . .	56
2.3.1	Maximal margin method . . . . .	56
2.3.1.1	Derivation of the maximal margin method . . . . .	56
2.3.1.2	Properties of the maximal margin method . . . . .	57
2.3.2	Linear $C$ -support vector classification . . . . .	59
2.3.2.1	Relationship between the primal and dual problems . . . . .	59
2.3.2.2	Linear $C$ -support vector classification . . . . .	60

<b>3</b>	<b>Linear Regression</b>	<b>63</b>
3.1	Regression Problems and Linear Regression Problems . . . . .	63
3.2	Hard $\bar{\varepsilon}$ -Band Hyperplane . . . . .	65
3.2.1	Linear regression problem and hard $\bar{\varepsilon}$ -band hyperplane	65
3.2.2	Hard $\bar{\varepsilon}$ -band hyperplane and linear classification . . . . .	66
3.2.3	Optimization problem of constructing a hard $\varepsilon$ -band hyperplane . . . . .	68
3.3	Linear Hard $\varepsilon$ -band Support Vector Regression . . . . .	70
3.3.1	Primal problem . . . . .	70
3.3.2	Dual problem and relationship between the primal and dual problems . . . . .	70
3.3.3	Linear hard $\varepsilon$ -band support vector regression . . . . .	74
3.4	Linear $\varepsilon$ -Support Vector Regression . . . . .	76
3.4.1	Primal problem . . . . .	76
3.4.2	Dual problem and relationship between the primal and dual problems . . . . .	77
3.4.3	Linear $\varepsilon$ -support vector regression . . . . .	79
<b>4</b>	<b>Kernels and Support Vector Machines</b>	<b>81</b>
4.1	From Linear Classification to Nonlinear Classification . . . . .	81
4.1.1	An example of nonlinear classification . . . . .	81
4.1.2	Classification machine based on nonlinear separation . . . . .	82
4.1.3	Regression machine based on nonlinear separation . . . . .	87
4.2	Kernels . . . . .	92
4.2.1	Properties . . . . .	93
4.2.2	Construction of kernels . . . . .	93
4.2.2.1	Basic kernels . . . . .	94
4.2.2.2	Operations keeping kernels . . . . .	94
4.2.2.3	Commonly used kernels . . . . .	96
4.2.2.4	Graph kernel . . . . .	97
4.3	Support Vector Machines and Their Properties . . . . .	101
4.3.1	Support vector classification . . . . .	101
4.3.1.1	Algorithm . . . . .	101
4.3.1.2	Support vector . . . . .	103
4.3.1.3	Properties . . . . .	104
4.3.1.4	Soft margin loss function . . . . .	105
4.3.1.5	Probabilistic outputs . . . . .	106
4.3.2	Support vector regression . . . . .	109
4.3.2.1	Algorithm . . . . .	109
4.3.2.2	Support vector . . . . .	110
4.3.2.3	Properties . . . . .	112
4.3.2.4	$\varepsilon$ -Insensitive loss function . . . . .	112
4.3.3	Flatness of support vector machines . . . . .	113
4.3.3.1	Runge phenomenon . . . . .	114
4.3.3.2	Flatness of $\varepsilon$ -support vector regression . . . . .	115

4.3.3.3 Flatness of  $C$ -support vector classification . . . 118

4.4 Meaning of Kernels . . . . . 120

**5 Basic Statistical Learning Theory of  $C$ -Support Vector Classification . . . . . 127**

5.1 Classification Problems on Statistical Learning Theory . . . 127

5.1.1 Probability distribution . . . . . 127

5.1.2 Description of classification problems . . . . . 131

5.2 Empirical Risk Minimization . . . . . 134

5.3 Vapnik Chervonenkis (VC) Dimension . . . . . 135

5.4 Structural Risk Minimization . . . . . 138

5.5 An Implementation of Structural Risk Minimization . . . . . 140

5.5.1 Primal problem . . . . . 140

5.5.2 Quasi-dual problem and relationship between quasi-dual problem and primal problem . . . . . 141

5.5.3 Structural risk minimization classification . . . . . 144

5.6 Theoretical Foundation of  $C$ -Support Vector Classification on Statistical Learning Theory . . . . . 145

5.6.1 Linear  $C$ -support vector classification . . . . . 145

5.6.2 Relationship between dual problem and quasi-dual problem . . . . . 146

5.6.3 Interpretation of  $C$ -support vector classification . . . . . 148

**6 Model Construction . . . . . 151**

6.1 Data Generation . . . . . 151

6.1.1 Orthogonal encoding . . . . . 152

6.1.2 Spectrum profile encoding . . . . . 153

6.1.3 Positional weighted matrix encoding . . . . . 154

6.2 Data Preprocessing . . . . . 155

6.2.1 Representation of nominal features . . . . . 155

6.2.2 Feature selection . . . . . 156

6.2.2.1  $F$ -score method . . . . . 156

6.2.2.2 Recursive feature elimination method . . . . . 157

6.2.2.3 Methods based on  $p$ -norm support vector classification ( $0 \leq p \leq 1$ ) . . . . . 158

6.2.3 Feature extraction . . . . . 164

6.2.3.1 Linear dimensionality reduction . . . . . 164

6.2.3.2 Nonlinear dimensionality reduction . . . . . 165

6.2.4 Data compression . . . . . 168

6.2.5 Data rebalancing . . . . . 169

6.3 Model Selection . . . . . 171

6.3.1 Algorithm evaluation . . . . . 171

6.3.1.1 Some evaluation measures for a decision function . . . . . 172

6.3.1.2	Some evaluation measures for a concrete algorithm . . . . .	175
6.3.2	Selection of kernels and parameters . . . . .	178
6.4	Rule Extraction . . . . .	180
6.4.1	A toy example . . . . .	180
6.4.2	Rule extraction . . . . .	182
<b>7</b>	<b>Implementation</b>	<b>187</b>
7.1	Stopping Criterion . . . . .	188
7.1.1	The first stopping criterion . . . . .	188
7.1.2	The second stopping criterion . . . . .	189
7.1.3	The third stopping criterion . . . . .	190
7.2	Chunking . . . . .	192
7.3	Decomposing . . . . .	194
7.4	Sequential Minimal Optimization . . . . .	197
7.4.1	Main steps . . . . .	198
7.4.2	Selecting the working set . . . . .	198
7.4.3	Analytical solution of the two-variables problem . . . . .	199
7.5	Software . . . . .	201
<b>8</b>	<b>Variants and Extensions of Support Vector Machines</b>	<b>203</b>
8.1	Variants of Binary Support Vector Classification . . . . .	203
8.1.1	Support vector classification with homogeneous decision function . . . . .	203
8.1.2	Bounded support vector classification . . . . .	206
8.1.3	Least squares support vector classification . . . . .	209
8.1.4	Proximal support vector classification . . . . .	211
8.1.5	$\nu$ -Support vector classification . . . . .	213
8.1.5.1	$\nu$ -Support vector classification . . . . .	213
8.1.5.2	Relationship between $\nu$ -SVC and $C$ -SVC . . . . .	215
8.1.5.3	Significance of the parameter $\nu$ . . . . .	215
8.1.6	Linear programming support vector classifications (LPSVC) . . . . .	216
8.1.6.1	LPSVC corresponding to $C$ -SVC . . . . .	216
8.1.6.2	LPSVC corresponding to $\nu$ -SVC . . . . .	218
8.1.7	Twin support vector classification . . . . .	218
8.2	Variants of Support Vector Regression . . . . .	224
8.2.1	Least squares support vector regression . . . . .	225
8.2.2	$\nu$ -Support vector regression . . . . .	226
8.2.2.1	$\nu$ -Support vector regression . . . . .	227
8.2.2.2	Relationship between $\nu$ -SVR and $\varepsilon$ -SVR . . . . .	229
8.2.2.3	The significance of the parameter $\nu$ . . . . .	229
8.2.2.4	Linear programming support vector regression (LPSVR) . . . . .	229
8.3	Multiclass Classification . . . . .	232

8.3.1	Approaches based on binary classifiers . . . . .	232
8.3.1.1	One versus one . . . . .	232
8.3.1.2	One versus the rest . . . . .	232
8.3.1.3	Error-correcting output coding . . . . .	234
8.3.2	Approach based on ordinal regression machines . . . . .	236
8.3.2.1	Ordinal regression machine . . . . .	237
8.3.2.2	Approach based on ordinal regression machines . . . . .	240
8.3.3	Crammer-Singer multiclass support vector classification . . . . .	243
8.3.3.1	Basic idea . . . . .	243
8.3.3.2	Primal problem . . . . .	243
8.3.3.3	Crammer-Singer support vector classification . . . . .	245
8.4	Semisupervised Classification . . . . .	247
8.4.1	PU classification problem . . . . .	247
8.4.2	Biased support vector classification <sup>[101]</sup> . . . . .	247
8.4.2.1	Optimization problem . . . . .	247
8.4.2.2	The selection of the parameters $C_+$ and $C_-$ . . . . .	249
8.4.3	Classification problem with labeled and unlabeled inputs . . . . .	250
8.4.4	Support vector classification by semidefinite programming . . . . .	251
8.4.4.1	Optimization problem . . . . .	251
8.4.4.2	Approximate solution via semidefinite programming . . . . .	252
8.4.4.3	Support vector classification by semidefinite programming . . . . .	254
8.5	Universum Classification . . . . .	255
8.5.1	Universum classification problem . . . . .	255
8.5.2	Primal problem and dual problem . . . . .	256
8.5.2.1	Algorithm and its relationship with three-class classification . . . . .	258
8.5.2.2	Construction of Universum . . . . .	258
8.6	Privileged Classification . . . . .	259
8.6.1	Linear privileged support vector classification . . . . .	259
8.6.2	Nonlinear privileged support vector classification . . . . .	261
8.6.3	A variation . . . . .	263
8.7	Knowledge-based Classification . . . . .	265
8.7.1	Knowledge-based linear support vector classification . . . . .	265
8.7.2	Knowledge-based nonlinear support vector classification . . . . .	268
8.8	Robust Classification . . . . .	272
8.8.1	Robust classification problem . . . . .	272
8.8.2	The solution when the input sets are polyhedrons . . . . .	273
8.8.2.1	Linear robust support vector classification . . . . .	273
8.8.2.2	Robust support vector classification . . . . .	277
8.8.3	The solution when the input sets are superspheres . . . . .	277

8.8.3.1	Linear robust support vector classification . . . . .	277
8.8.3.2	Robust support vector classification . . . . .	281
8.9	Multi-instance Classification . . . . .	282
8.9.1	Multi-instance classification problem . . . . .	282
8.9.2	Multi-instance linear support vector classification . . . . .	283
8.9.2.1	Optimization problem . . . . .	283
8.9.2.2	Linear support vector classification . . . . .	286
8.9.3	Multi-instance support vector classification . . . . .	288
8.10	Multi-label Classification . . . . .	292
8.10.1	Problem transformation methods . . . . .	292
8.10.2	Algorithm adaptation methods . . . . .	294
8.10.2.1	A ranking system . . . . .	294
8.10.2.2	Label set size prediction . . . . .	296
8.10.3	Algorithm . . . . .	297
	<b>Bibliography</b>	<b>299</b>
	<b>Index</b>	<b>315</b>