

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	3
1.2	Zielsetzung und Untersuchungsgegenstand . . . . .	7
1.3	Vorschau . . . . .	9
1.3.1	Übersicht Kapitel 2–12 . . . . .	9
1.3.2	Hinweise zur Implementation . . . . .	12
<b>2</b>	<b>Mathematische Grundlagen</b>	<b>13</b>
2.1	Allgemeine Bezeichnungen I . . . . .	14
2.2	Vereinfachte Bezeichnungen im $\mathbb{G}^d$ . . . . .	15
2.3	Normalteilungen im $\mathbb{R}^d$ . . . . .	18
2.4	Maß- und Wahrscheinlichkeitstheorie . . . . .	20
2.4.1	Motivation . . . . .	20
2.4.2	Grundlagen . . . . .	22
2.4.3	Lebesgue-Stieltjes-Wahrscheinlichkeitsmaß . . . . .	38
2.4.4	Bedingte Wahrscheinlichkeiten . . . . .	55
2.4.5	Momente . . . . .	58
2.4.6	Grenzwertsätze . . . . .	61
2.5	Allgemeine Bezeichnungen II . . . . .	66
2.6	Graphentheorie . . . . .	67
2.7	Ergänzungen . . . . .	71
<b>3</b>	<b>Einführung in die Thematik</b>	<b>73</b>
3.1	QSAR-Modelle . . . . .	73
3.1.1	Strukturraum . . . . .	74
3.1.2	Deskriptorraum . . . . .	76
3.1.3	Anwendungsdomäne . . . . .	78
3.2	Over-/ Underfitting . . . . .	85

<b>4 Konventionelle AD-Schätzer</b>	<b>87</b>
4.1 Überblick . . . . .	87
4.1.1 Bereichsbezogene und geometrische Methode . . . . .	87
4.1.2 Distanzbasierte Methoden . . . . .	89
4.2 Mahalanobis-Norm . . . . .	95
4.2.1 Formale Herleitung . . . . .	98
4.3 Leverage . . . . .	109
<b>5 Nichtparametrische Kerndichteschätzung</b>	<b>115</b>
5.1 Ursprung . . . . .	116
5.2 Grundlegende Begriffe . . . . .	118
5.2.1 Der univariate Fall . . . . .	118
5.2.2 Der multivariate Fall . . . . .	124
5.3 Bandbreitenwahl . . . . .	129
5.3.1 Optimalitätskriterien . . . . .	130
5.3.2 Kreuzvalidierung der kleinsten Quadrate . . . . .	130
5.3.3 Beurteilung der Kreuzvalidierungsverfahren . . . . .	132
<b>6 Der kernbasierte AD-Schätzer KADE</b>	<b>133</b>
6.1 Datenaufbereitung und Skalierung . . . . .	133
6.1.1 Whitening-Transformation . . . . .	134
6.2 KADE-Standardbandbreite . . . . .	139
6.2.1 Anfragegesteuerte Bandbreite . . . . .	144
6.3 Donänenbegrenzung . . . . .	145
6.4 Zusammenfassung . . . . .	146
<b>7 Datenstrukturen</b>	<b>147</b>
7.1 Raumteilende Bäume . . . . .	148
7.1.1 k-d-Bäume . . . . .	149
7.1.2 Metrische Bäume . . . . .	150
7.1.3 Entartungen . . . . .	156
7.2 Anker-Hierarchie . . . . .	167
7.2.1 Motivation . . . . .	167
7.2.2 Aufbau . . . . .	169

---

<b>8 HDR-Berechnung</b>	<b>175</b>
8.1 Motivation	175
8.1.1 Grundlagen	176
8.2 Monte-Carlo-Integration	180
8.2.1 Methode der wesentlichen Stichprobe	182
8.2.2 Die wesentliche Stichprobe bei Kerndichteschätzern	184
8.2.3 Geschichtete Zufallszahlen	190
8.3 Monte-Carlo-HDR-Schätzer	191
8.3.1 Integralwertkorrigierter Monte-Carlo-HDR-Schätzer	194
8.3.2 Bezug zum KADE-AD-Cutoff	195
<b>9 Der zielraumgestützte AD-Schätzer EKADE</b>	<b>197</b>
9.1 Vorschläge aus der Literatur	197
9.2 Berücksichtigung des Modellfehlers	201
9.2.1 Die Fehlergewichtsfunktion	203
9.2.2 Domänenbegrenzung	206
9.3 Erweiterung des Basisdatensatzes	210
<b>10 Optimalitätskriterien für AD-Schätzer</b>	<b>215</b>
10.1 Verwendung konventioneller Maße	216
10.2 Das $\aleph$ -Maß	218
<b>11 Vergleichsstudie</b>	<b>231</b>
11.1 Untersuchte QSAR-Modelle	231
11.2 Methodik	236
11.3 Ergebnisse und Diskussion	237
11.3.1 Studienteil A: KADE	237
11.3.2 Studienteil B: EKADE	254
<b>12 Erweiterte Anwendungen</b>	<b>271</b>
<b>13 Schlussbemerkungen und Ausblick</b>	<b>277</b>

<b>Anhang</b>	<b>283</b>
<b>A Ergebnisse Beispielrechnungen</b>	<b>283</b>
A.1 Studien zu Kapitel 6 . . . . .	283
A.1.1 Beispiel 6.2.1 . . . . .	283
A.2 Studien zu Kapitel 7 . . . . .	284
A.2.1 Beispiel 7.1.2, $b=64$ . . . . .	284
A.2.2 Beispiel 7.1.2, $b=256$ . . . . .	290
A.3 Studien zu Kapitel 8 . . . . .	308
A.3.1 Beispiele 8.2.1 bis 8.2.4 . . . . .	308
A.3.2 Beispiele 8.3.1 bis 8.3.2 . . . . .	330
A.4 Studien zu Kapitel 9 . . . . .	353
A.4.1 Beispiel 9.2.1 . . . . .	353
<b>B Ergänzende Informationen</b>	<b>355</b>
B.1 Anmerkungen zu Kapitel 10 . . . . .	355
B.2 Anmerkungen zu Kapitel 11 . . . . .	356
B.2.1 Regressionsgewichtung . . . . .	356
B.2.2 Notation der Deskriptoren . . . . .	357
B.2.3 Zusammenhang zwischen den Zielwerten . . . . .	357
<b>C Literaturmodelle</b>	<b>359</b>
C.1 Modell M1 . . . . .	359
C.1.1 Trainingsdaten . . . . .	359
C.1.2 Validierungsdaten . . . . .	368
C.2 Modell M2 . . . . .	376
C.2.1 Trainingsdaten . . . . .	376
C.2.2 Validierungsdaten . . . . .	387
C.3 Modell M3 . . . . .	407
C.3.1 Trainingsdaten . . . . .	407
C.3.2 Validierungsdaten . . . . .	422
C.4 Modell M4 . . . . .	425
C.4.1 Trainingsdaten . . . . .	426
C.4.2 Validierungsdaten . . . . .	430

---

C.5	Modell M5 . . . . .	449
C.5.1	Trainingsdaten . . . . .	449
C.5.2	Validierungsdaten . . . . .	452
C.6	Modell M6 . . . . .	455
C.6.1	Trainingsdaten . . . . .	456
C.6.2	Validierungsdaten . . . . .	457
C.7	Modell M7 . . . . .	462
C.7.1	Trainingsdaten . . . . .	462
C.7.2	Validierungsdaten . . . . .	468
<b>Literaturverzeichnis</b>		<b>491</b>
<b>Liste verwendeter Symbole</b>		<b>509</b>
<b>Register</b>		<b>517</b>