

Inhaltsverzeichnis

1	Einleitung: Vom Batch Machine Learning zum Online Machine Learning	1
	Thomas Bartz-Beielstein	
1.1	Datenströme	1
1.2	Nachteile des Batch-Lernens	3
1.3	Inkrementelles Lernen, Online-Lernen und Stream-Lernen	11
1.4	Überführung des Batch- in das Online Machine-Learning	13
	Literaturverzeichnis	13
2	Supervised Learning: Klassifikation und Regression	15
	Thomas Bartz-Beielstein	
2.1	Klassifikation	15
2.2	Regression	21
2.3	Ensemble-Methoden für OML	22
2.4	Clustering	23
2.5	Übersicht: OML-Verfahren	23
	Literaturverzeichnis	24
3	Drifterkennung und -behandlung	25
	Thomas Bartz-Beielstein	
3.1	Architekturen für Driftbehandlungsmethoden	25
3.2	Grundlegende Überlegungen zu Fenstertechniken	27
3.3	Populäre Verfahren zur Drifterkennung	28
3.4	OML-Algorithmen mit Drifterkennung: Hoeffding-Window Trees	32
	Literaturverzeichnis	37
4	Initiale Auswahl und nachträgliche Aktualisierung von OML Modellen	41
	Thomas Bartz-Beielstein	
4.1	Initiale Modellauswahl	41
4.2	Modelländerungen	43
4.3	Katastrophales Vergessen	44

	Literaturverzeichnis	46
5	Evaluation und Performance-Messung	47
	Thomas Bartz-Beielstein	
5.1	Auswahlmethode	47
5.2	Bestimmung des Training- und Testdatensatzes im Paket spotRiver	54
5.3	Performanz des Algorithmus/Modells	61
5.4	Datenstrom- und Drift-Generatoren	63
5.5	Zusammenfassung	65
	Literaturverzeichnis	65
6	Besondere Anforderungen an OML-Verfahren	67
	Thomas Bartz-Beielstein	
6.1	Fehlende Daten, Imputation	67
6.2	Kategorische Attribute	68
6.3	Ausreißer	69
6.4	Unbalancierte (unausgewogene) Daten	70
6.5	Große Anzahl an Features (Attributen)	71
6.6	FAIR, Interpretierbarkeit und Erklärbarkeit	71
	Literaturverzeichnis	73
7	Praxisanwendungen	75
	Steffen Moritz, Florian Dumpert, Thomas Bartz-Beielstein, Eva Bartz	
7.1	Anwendungen und Anwendungsperspektiven in der amtlichen Statistik	76
	Exkurs: Nowcasting	80
7.2	Andere Anwendungen mit Bezug zur amtlichen Statistik	86
7.3	Aspekte bezüglich des Praxiseinsatzes	90
7.4	Zusammenfassung	92
	Literaturverzeichnis	93
8	Open-Source-Software für Online Machine Learning	95
	Thomas Bartz-Beielstein	
8.1	Übersicht und Beschreibung der Softwarepakete	95
8.2	Softwareumfang	99
8.3	Vergleich der Programmiersprachen	101
	Literaturverzeichnis	102
9	Ein experimenteller Vergleich von Batch- und Online Machine Learning-Algorithmen	103
	Thomas Bartz-Beielstein	
9.1	Studie: Bike Sharing	103
9.2	Studie: Sehr große Datensätze mit Drift	117
9.3	Zusammenfassung	120
	Literaturverzeichnis	122

- 10 Hyperparameter Tuning** 123
 - Thomas Bartz-Beielstein
 - 10.1 Hyperparameter Tuning: Eine Einführung 123
 - 10.2 Die Hyperparameter-Tuning-Software SPOT 124
 - 10.3 Studie: Hyperparameter Tuning des HTR-Algorithmus auf den
Friedman-Drift-Daten 126
 - River-spezifisches Setup 129
 - 10.4 Vergleich der HTR- und HATR-Algorithmen 139
 - 10.5 Verwendung der spotPython GUI 144
 - 10.6 Zusammenfassung 146
 - 10.7 Übersicht der experimentellen Setups. 148
 - Literaturverzeichnis 150

- 11 Zusammenfassung und Ausblick** 151
 - Thomas Bartz-Beielstein, Eva Bartz
 - 11.1 Notwendigkeit für OML-Verfahren 151
 - 11.2 Empfehlungen für die OML-Praxis 152
 - Literaturverzeichnis 154

- A Definitionen und Erläuterungen** 155
 - A.1 Gradientenabstieg 155
 - A.2 Satz von Bayes 155
 - A.3 Hoeffding Schranke 156
 - A.4 Kappa Statistiken 156

- B Zusatzmaterial** 159
 - B.1 Notebooks 159
 - B.2 Software 160

- Sachverzeichnis** 165