

Wissensentdeckung in Datenbanken

Übersicht, Literaturhinweise

Nico Piatkowski und Uwe Ligges

Informatik—Künstliche Intelligenz
Computergestützte Statistik
Technische Universität Dortmund

27.07.2017



Übersicht, Literatur

- Grundbegriffe
 - Daten, Modell, Verlustfunktion, Lernen, Vorhersage

The Elements of Statistical Learning; 2.4

- Modellklassen, Verlustfunktionen, Parameterschätzung
 - Gradientenabstieg, Eigenschaften von Verlustfunktionen

Convex Optimization; 3.1–3.1.3, 9.3

- Überanpassung, Regularisierung

The Elements of Statistical Learning; 3.4.1, 3.4.2, 7.1, 7.2, 7.3



Übersicht, Literatur

- Grundbegriffe
 - Daten, Modell, Verlustfunktion, Lernen, Vorhersage

The Elements of Statistical Learning; 2.4

- Modellklassen, Verlustfunktionen, Parameterschätzung
 - Gradientenabstieg, Eigenschaften von Verlustfunktionen

Convex Optimization; 3.1–3.1.3, 9.3

- Überanpassung, Regularisierung

The Elements of Statistical Learning; 3.4.1, 3.4.2, 7.1, 7.2, 7.3



Übersicht, Literatur

- Grundbegriffe
 - Daten, Modell, Verlustfunktion, Lernen, Vorhersage

The Elements of Statistical Learning; 2.4

- Modellklassen, Verlustfunktionen, Parameterschätzung
 - Gradientenabstieg, Eigenschaften von Verlustfunktionen

Convex Optimization; 3.1–3.1.3, 9.3

- Überanpassung, Regularisierung

The Elements of Statistical Learning; 3.4.1, 3.4.2, 7.1, 7.2, 7.3



Übersicht, Literatur

- Datenbanken, SQL

https://www.sqlite.org/lang_select.html

- Häufige Mengen, Apriori, FP-growth

Artikel zu FPGrowth

- SVM

- Lineare Trennbarkeit, Hyperebenen
- Merkmalstransformationen (ϕ), Lagrange-Dual, Kernel

The Elements of Statistical Learning; 12.2, 12.3.1, 12.3.2



Übersicht, Literatur

- Datenbanken, SQL

https://www.sqlite.org/lang_select.html

- Häufige Mengen, Apriori, FP-growth

Artikel zu FPGrowth

- SVM

- Lineare Trennbarkeit, Hyperebenen
- Merkmalstransformationen (ϕ), Lagrange-Dual, Kernel

The Elements of Statistical Learning; 12.2, 12.3.1, 12.3.2



Übersicht, Literatur

- Datenbanken, SQL

https://www.sqlite.org/lang_select.html

- Häufige Mengen, Apriori, FP-growth

Artikel zu FPGrowth

- SVM
 - Lineare Trennbarkeit, Hyperebenen
 - Merkmalstransformationen (ϕ), Lagrange-Dual, Kernel

The Elements of Statistical Learning; 12.2, 12.3.1, 12.3.2

Übersicht, Literatur

- Stochastische Abhängigkeit vs. Korrelation
- Chow-Liu Bäume, Abhängigkeitsgraphen mittels Regularisierung
- Probabilistische Graphische Modelle
 - Markov Eigenschaft, Faktorisierung der Wahrscheinlichkeitsdichte
 - Graphen, Suffiziente Statistik, Maximum Entropy
 - Belief Propagation

Graphical Models, Exponential Families, and Variational Inference; 2.5.1, 3.1, 3.2, 3.3

- Proximaler Gradientenabstieg

Proximal Algorithms; 1.1–1.3, 4.2, 6.1.3

Übersicht, Literatur

- Stochastische Abhängigkeit vs. Korrelation
- Chow-Liu Bäume, Abhängigkeitsgraphen mittels Regularisierung
- Probabilistische Graphische Modelle
 - Markov Eigenschaft, Faktorisierung der Wahrscheinlichkeitsdichte
 - Graphen, Suffiziente Statistik, Maximum Entropy
 - Belief Propagation

Graphical Models, Exponential Families, and Variational Inference; 2.5.1, 3.1, 3.2, 3.3

- Proximaler Gradientenabstieg

Proximal Algorithms; 1.1–1.3, 4.2, 6.1.3



Übersicht, Literatur

- Stochastische Abhängigkeit vs. Korrelation
- Chow-Liu Bäume, Abhängigkeitsgraphen mittels Regularisierung
- Probabilistische Graphische Modelle
 - Markov Eigenschaft, Faktorisierung der Wahrscheinlichkeitsdichte
 - Graphen, Suffiziente Statistik, Maximum Entropy
 - Belief Propagation

Graphical Models, Exponential Families, and Variational Inference; 2.5.1, 3.1, 3.2, 3.3

- Proximaler Gradientenabstieg

Proximal Algorithms; 1.1–1.3, 4.2, 6.1.3



Übersicht, Literatur

- Merkmalsauswahl, Greedy Selection, Regularisierung

The Elements of Statistical Learning; 3.3.1–3.3.3, 3.4.2

- Darstellung von Textdaten (Bag-of-Words, TF-IDF)
- Clustering
 - k -means, Optimalität der Clusteranzahl, DBSCAN, Latent Dirichlet Allocation

The Elements of Statistical Learning; 14.3.6, 14.3.7
Graphical Models, Exponential Families, and Variational Inference; Example 3.5



Übersicht, Literatur

- Merkmalsauswahl, Greedy Selection, Regularisierung

The Elements of Statistical Learning; 3.3.1–3.3.3, 3.4.2

- Darstellung von Textdaten (Bag-of-Words, TF-IDF)
- Clustering
 - k -means, Optimalität der Clusteranzahl, DBSCAN, Latent Dirichlet Allocation

The Elements of Statistical Learning; 14.3.6, 14.3.7

Graphical Models, Exponential Families, and Variational Inference; Example 3.5



Übersicht, Literatur

- Deep Learning
 - Künstliche Neuronale Netze, Aufbau, Aktivierungsfunktionen, Parameterlernen
 - Faltungsnetze
 - Dropout, Autoencoder
 - Generative Adversarial Netze

Artikel zu Deep Convolutional Neural Networks

Artikel zu Dropout

Artikel zu GANs

Inhalt

- 18.4. L,P Übersicht, Einführung
- 20.4. L Statistik 1
- 25.4. L Statistik 2
- 27.4. L Stichproben, Versuchsplanung, Datenvorverarbeitung
- 02.5. Optimierung, Modellklassen, Lineares Modell,
P Bias-Varianz, Overfitting 1
- 04.5. Optimierung, Modellklassen, Lineares Modell,
P Bias-Varianz, Overfitting 2
- 09.5. P Data Cube, Frequent sets, Apriori, FPgrowth 1
- 11.5. P Data Cube, Frequent sets, Apriori, FPgrowth 2
- 16.5. L Einführung Klassifikation, BayesRegel,
Logistische Regression
- 18.5. L kNN, Ähnlichkeitsmaße, Modellselektion
- 23.5. L Resampling, Klassifikationsbeurteilung 2

Inhalt

- 30.5. P SVM
- 01.6. L Diskriminanzanalyse (LDA) 1
- 06.6. L Diskriminanzanalyse (LDA, QDA, RDA) 2
- 08.6. L von Entscheidungsbäumen (CART) zu Wäldern
- 13.6. L Ensemble Methoden (Bagging, Boosting)
- 20.6. L Stetige Modelle Hauptkomponentenanalyse)
- 22.6. P Graphische Modelle 1
- 27.6. P Graphische Modelle 2
- 29.6. P Nicht-glatte und stochastische Optimierung
- 04.7. P Merkmalsselektion, Struktur Lernen, Regularisierung

Inhalt

- 06.7. P Clustering, k-Means, Gaussian Mixture, Latent Dirichlet Allocation 1
- 11.7. P Clustering, k-Means, Gaussian Mixture, Latent Dirichlet Allocation 2
- 13.7. L Hierarchisches Clustern; Zeitreihen 1
- 18.7. L Zeitreihen 2
- 20.7. P Künstliche Neuronale Netze 1
- 25.7. P Künstliche Neuronale Netze 2
- 27.7. L,P Zusammenfassung; Rückblick