

Klassifikation funktionaler Daten

Präsentation Masterseminar

Betreuerin: Prof. Dr. Sonja Greven

Thomas Maierhofer

Institut für Statistik
Ludwig-Maximilians-Universität München

27. Januar 2017



www.email-lounge.de/allgemein/studien

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten
 - Nonparametric Functional Kernel Estimator
 - Nächste Nachbarn Ensemble
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

Motivation: Klassifikation funktionaler Daten

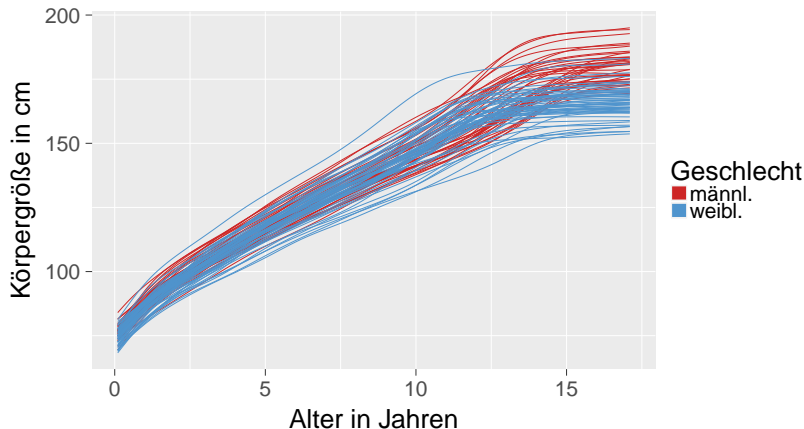


Abbildung: Daten der Berkely Growth Study (Tuddenham and Snyder; 1954). Darstellung der Wachstumskurven der Jungen (rot) und Mädchen (blau) über den Beobachtungszeitraum, Mittelwertskurven in schwarz.

Motivation: Klassifikation funktionaler Daten

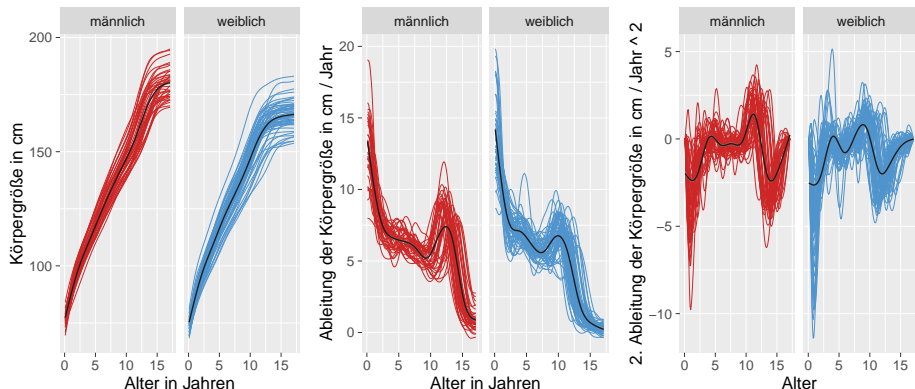


Abbildung: Darstellung der Wachstumskurven und deren Ableitungen der Jungen (rot) und Mädchen (blau) mit Mittelwertskurven in schwarz.

1 Motivation

2 Zentrale Konzepte

- Funktionale Datenanalyse
- Klassifikation
- Semimetriken

3 Methoden zur Klassifikation funktionaler Daten

- Nonparametric Functional Kernel Estimator
- Nächste Nachbarn Ensemble
- Kontrastierung und Einordnung der beiden Methoden

4 Vergleich der Methoden

5 Zusammenfassung & Diskussion

Funktionale Datenanalyse (engl. Functional Data Analysis, FDA)

- Kurven, Oberflächen und sämtlichen Daten, die auf einem kontinuierlichen Träger (zB. Zeit, Raum, Wellenlänge, Winkel, ...) erhoben werden
- In realen Datensätzen ist zugrundeliegende Funktion nur an diskreter Menge von Punkten beobachtet

Klassifikation

- Zuordnung neuer Beobachtungen mit unbekannter Klasse zu bekannten Klassen
- Verwende bereits bekannte Beobachtungen mit bekannter Klasse um Zuordnungsvorschriften zu erstellen

1 Motivation

2 Zentrale Konzepte

- Funktionale Datenanalyse
- Klassifikation
- Semimetriken

3 Methoden zur Klassifikation funktionaler Daten

- Nonparametric Functional Kernel Estimator
- Nächste Nachbarn Ensemble
- Kontrastierung und Einordnung der beiden Methoden

4 Vergleich der Methoden

5 Zusammenfassung & Diskussion

Abgrenzung zur multivariaten Statistik

- Multivariate Statistik behandelt m -dimensionale Zufallsvariablen, die keine natürliche Ordnung haben
- FDA behandelt ∞ -dimensionale Zufallsvariablen $x(t)$, $t \in \mathcal{T}$, \mathcal{T} Intervall auf \mathbb{R} (oder Bereich auf \mathbb{R}^n), die an beliebig vielen Punkten auswertbar sind

Abgrenzung zur Zeitreihenanalyse

- Zeitreihen sind einmalige nicht-wiederholbare Verläufe
- Funktionale Daten sind (wiederholbare) Realisation einer funktionalen Zufallsvariable
- Eine Beobachtung ist in der Zeitreihenanalyse ein Messzeitpunkt, in FDA eine ganze Kurve
- In FDA werden die Daten meist als vollständig bekannt gesehen, Vorhersage des weiteren Kurvenverlaufs eher in Zeitreihenanalyse

1 Motivation

2 Zentrale Konzepte

- Funktionale Datenanalyse
- **Klassifikation**
- Semimetriken

3 Methoden zur Klassifikation funktionaler Daten

- Nonparametric Functional Kernel Estimator
- Nächste Nachbarn Ensemble
- Kontrastierung und Einordnung der beiden Methoden

4 Vergleich der Methoden

5 Zusammenfassung & Diskussion

Definition Klassifikation

- Methoden und Kriterien zur Einteilung von Objekten in Klassen
- Zuordnung neuer Beobachtungen mit unbekannter Klasse zu bekannten Klassen

Abgrenzung zur Clusteranalyse

- Clusteranalyse findet Regeln um Beobachtungen zu gruppieren, Klassifikation findet Regeln um Gruppen/Klassen zu unterscheiden
- Klassen müssen in Clusteranalyse gefunden werden (Unsupervised Learning), sind in Klassifikation bereits bekannt oder natürlich gegeben (Supervised Learning)

1 Motivation

2 Zentrale Konzepte

- Funktionale Datenanalyse
- Klassifikation
- Semimetriken

3 Methoden zur Klassifikation funktionaler Daten

- Nonparametric Functional Kernel Estimator
- Nächste Nachbarn Ensemble
- Kontrastierung und Einordnung der beiden Methoden

4 Vergleich der Methoden

5 Zusammenfassung & Diskussion

Eine Funktion d ist eine Semimetrik auf einem Raum \mathcal{X} wenn:

- $\forall x \in \mathcal{X} : d(x, x) = 0,$
- $\forall x_1, x_2, x_3 \in \mathcal{X} : d(x_1, x_2) \leq d(x_1, x_3) + d(x_3, x_2)$

Für eine Metrik wäre die zusätzliche Forderung notwendig, dass
 $\forall x_1, x_2 \in \mathcal{X} : d(x_1, x_2) = 0 \Rightarrow x_1 = x_2.$

Beispiel: Euklidische Distanz der Ableitungen

Die euklidische Distanz $d_a^{\text{Eukl.}}(\cdot, \cdot)$ der Ableitung a -ten Grades zweier Beobachtungen $x_1(t)$ und $x_2(t)$ ist definiert als

$$d_a^{\text{Eukl.}}(x_1(t), x_2(t)) = \sqrt{\int_{\mathcal{T}} \left(x_1^{(a)}(t) - x_2^{(a)}(t) \right)^2 dt}.$$

- Für $a = 0$, handelt es sich hierbei um eine Metrik
- Für $a > 0$ lediglich Semimetrik, da $d_a^{\text{Eukl.}}(x_1(t), x_2(t)) = 0$ für $x_2(t) = x_1(t) + c$
- Einschränkung bzw. Gewichtung des Integrationsbereichs ebenfalls möglich

Beispiel: Euklidische Distanz der Ableitungen

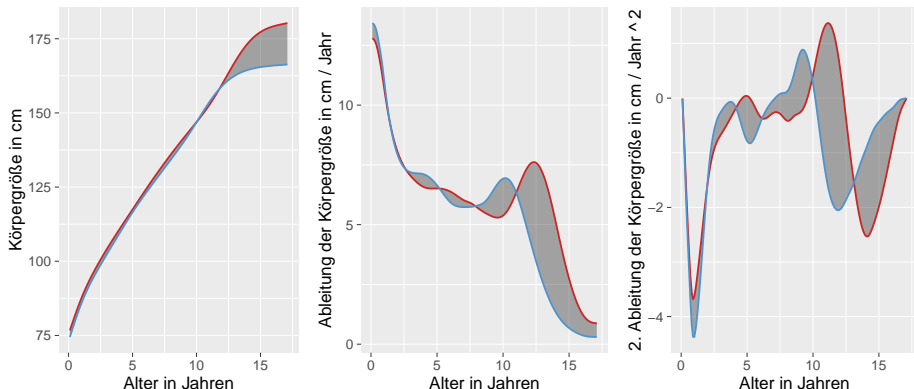


Abbildung: Darstellung der Mittelwertskurven der Jungen (rot) und Mädchen (blau) über den Beobachtungszeitraum.

Beispiel: Differenz der globalen Maxima/Minima der Ableitungen

Die Differenz der globalen Minima/Maxima der Ableitung vom Grad a bzw. der Originalkurve (für $a = 0$) zweier Beobachtungen $x_1(t)$ und $x_2(t)$ ist eine Semimetrik:

$$d_a^{\text{Max.}}(x_1(t), x_2(t)) = |\max_t(x_1^{(a)}(t)) - \max_t(x_2^{(a)}(t))|$$

$$d_a^{\text{Min.}}(x_1(t), x_2(t)) = |\min_t(x_1^{(a)}(t)) - \min_t(x_2^{(a)}(t))|$$

Beispiel: Differenz der globalen Maxima/Minima der Ableitungen

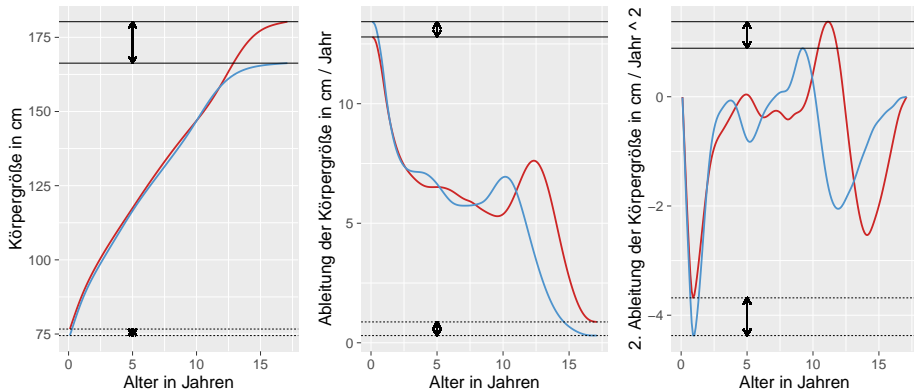


Abbildung: Mittelwertskurven der Jungen (rot) und Mädchen (blau) und deren Minima (gestrichelte Linie) bzw. Maxima (durchgezogene Linie).

1 Motivation

2 Zentrale Konzepte

- Funktionale Datenanalyse
- Klassifikation
- Semimetriken

3 Methoden zur Klassifikation funktionaler Daten

- Nonparametric Functional Kernel Estimator
- Nächste Nachbarn Ensemble
- Kontrastierung und Einordnung der beiden Methoden

4 Vergleich der Methoden

5 Zusammenfassung & Diskussion

Nonparametric Functional Kernel Estimator

- Aus Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach, Computational Statistics & Data Analysis 44(1): 161–173.
- Basiert auf Kerndichte Schätzer

Nächste Nachbarn Ensemble

- Aus Fuchs, K., Gertheiss, J. and Tutz, G. (2015). Nearest neighbor ensembles for functional data with interpretable feature selection, Chemometrics and Intelligent Laboratory Systems 146: 186–197.
- Basiert auf Nächste Nachbarn Schätzern

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten**
 - Nonparametric Functional Kernel Estimator**
 - Nächste Nachbarn Ensemble
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

Nonparametric Functional Kernel Estimator

Geschätzte Wahrscheinlichkeit $\hat{\pi}_{g,h}$ der Zugehörigkeit einer Beobachtung $x^*(t)$ zu einer Klasse $g \in G$ ist

$$\hat{\pi}_{g,h}(x^*(t)) = \frac{\sum_{j=1}^N I(y_j = g) K(d(x^*(t), x_j(t))/h)}{\sum_{j=1}^N K(d(x^*(t), x_j(t))/h)}.$$

mit

- $x_j(t), j = 1, \dots, N$ den Beobachtungen im Trainingsdatensatz
- der Indikatorfunktion $I(g_1 = g_2) = 1 \Leftrightarrow g_1 = g_2$, sonst 0

und im Voraus festzulegender

- Semimetrik $d(\cdot, \cdot)$
- Kernfunktion $K(\cdot)$
- Bandweite h

Prognostiziere Klasse mit größter Wahrscheinlichkeit $\underset{g \in G}{\operatorname{argmax}} \hat{\pi}_{g,h}(x^*(t)).$

Beispiel: Nonparametric Functional Kernel Estimator

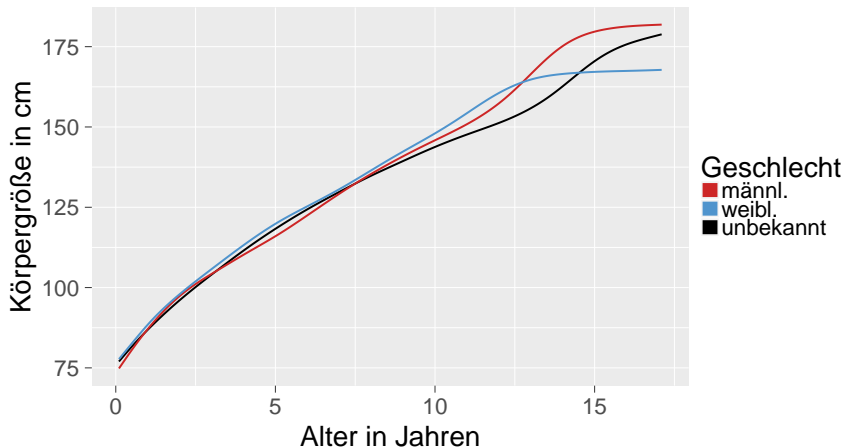


Abbildung: Wachstumskurven eines Jungen (rot), eines Mädchens (blau) und eines Kindes unbekannten Geschlechts (schwarz) aus der Berkely Growth Study.

Beispiel: Nonparametric Functional Kernel Estimator

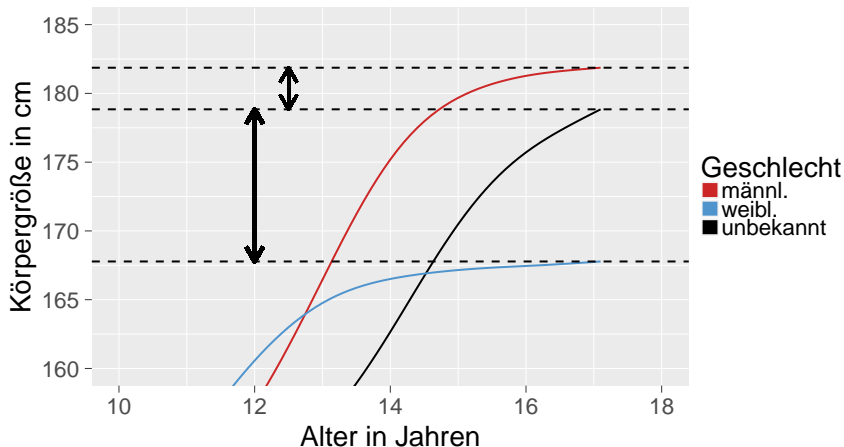


Abbildung: Ausschnitt aus Abbildung 5. Zusätzlich sind die Maxima der Wachstumskurven eingetragen und die Differenzen zur neuen Beobachtung.

Beispiel: Nonparametric Functional Kernel Estimator

Schätzgleichung der Klassenzugehörigkeit:

$$\hat{\pi}_{g,h}(x^*(t)) = \frac{\sum_{j=1}^N I(y_j = g) K(d(x^*(t), x_j(t))/h)}{\sum_{j=1}^N K(d(x^*(t), x_j(t))/h)}$$

mit

- Kernfunktion $K(u) = (1 - |u|) I(|u| \leq 1)$ (Dreieckskern)
- Maximumsdistanz $d^{\text{Max.}}(x_1(t), x_2(t)) = |\max_t(x_1(t)) - \max_t(x_2(t))|$
- Bandweite $h = 20$

Hier

- $d^{\text{Max.}}(x_m(t), x_{un}(t)) = |182 - 179| = 3$
- $d^{\text{Max.}}(x_w(t), x_{un}(t)) = |168 - 179| = 11$
- $\hat{\pi}_{m,20}(x_{un}(t)) = \frac{K(3/20)}{K(3/20) + K(11/20)} = 0.65$
- $\hat{\pi}_{w,20}(x_{un}(t)) = \frac{K(11/20)}{K(3/20) + K(11/20)} = 0.35$

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten**
 - Nonparametric Functional Kernel Estimator
 - Nächste Nachbarn Ensemble**
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

Zusatzanforderung an Semimetrik $d(\cdot, \cdot)$: Symmetrie

Zusätzlich zu den in Abschnitt 3 genannten Eigenschaften, dass $\forall x_1, x_2 \in \mathcal{X} : d(x_1, x_2) = d(x_2, x_1)$.

Einführung der Notation

- Sei $(y, x(t))$ Beobachtung mit Klasse $y_i \in G$ und funktionaler Kovariable $x(t)$
- Seien $(y_i, x_i(t)), i = 1, \dots, N$ die Beobachtungen im Trainingsdatensatz mit bekannter Klassenzugehörigkeit
- $(y^*, x^*(t))$ eine Neue zu klassifizierende Beobachtung

Beispiel: Nächste Nachbarn Klassifikation

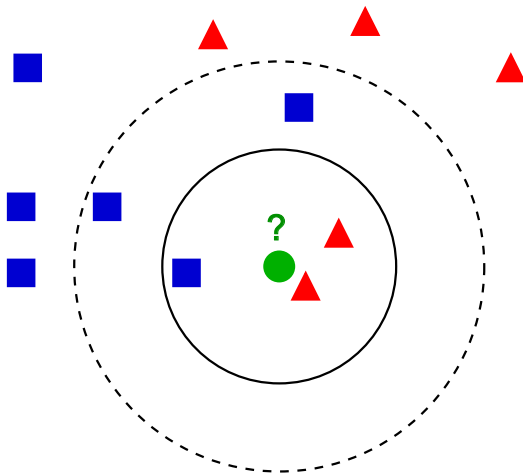


Abbildung: Schematische Darstellung der Nächste Nachbarn Klassifikation. Für $k = 3$ wird die Klasse „rote Dreiecke“ geschätzt, für $k = 5$ „blaue Quadrate“. (https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)

Definition Nachbarschaft

Sortiere $x_i(t)$, $i = 1, \dots, N$ nach Abstand zu $x^*(t)$ bezüglich $d(\cdot, \cdot)$, also

$$d(x^*(t), x_{(1)}(t)) \leq \dots \leq d(x^*(t), x_{(k)}(t)) \leq \dots \leq d(x^*(t), x_{(N)}(t)).$$

Nachbarschaft $\mathcal{N}(x^*(t))$ der k nächsten Nachbarn von $x^*(t)$ ist

$$\mathcal{N}(x^*(t)) = \{x_j(t) : d(x^*(t), x_j(t)) \leq d(x^*(t), x_{(k)}(t))\}.$$

Nächste Nachbarn Schätzer für Klassenzugehörigkeit

Schätzer für Zugehörigkeit von x^* zu Klasse $g \in G$ ist

$$\hat{\pi}_g = \frac{1}{k} \sum_{x_j(t) \in \mathcal{N}(x^*(t))} I(y_j = g).$$

Schätzer für y^* ist die Klasse mit höchster Wahrscheinlichkeit

$$\hat{y}^* = \underset{g \in G}{\operatorname{argmax}} (\hat{\pi}_g).$$

Erweiterung auf mehrere Semimetriken

- Verwende p Semimetriken $d_l(\cdot, \cdot)$, $l = 1, \dots, p$ statt einer einzigen Semimetrik $d(\cdot, \cdot)$
- Bestimme jeweils die eigene Nachbarschaft $\mathcal{N}_l(x^*(t))$
- Nächste Nachbarn Schätzer für jede Semimetrik ist

$$\hat{\pi}_{g,l} = \frac{1}{k} \sum_{x_j(t) \in \mathcal{N}_l(x^*(t))} I(y_j = g)$$

Ensemble Schätzer

Ensemble Schätzer ist gewichtete Summe der einzelnen Schätzungen,

$$\hat{\pi}_g = \sum_{l=1}^p c_l \hat{\pi}_{gl},$$

wobei Gewichte c_l den Restriktionen $c_l \geq 0 \forall l$ und $\sum_{l=1}^p c_l = 1$ unterliegen und Brier Score-optimal ausgewählt werden.

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten**
 - Nonparametric Functional Kernel Estimator
 - Nächste Nachbarn Ensemble
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

Gemeinsamkeiten

- Ähnlichkeit zwischen Beobachtungen über Semimetriken definiert
- Klassifikation einer neuen Beobachtung basiert auf (gewichtetem) Auszählen der Klassen der ähnlichsten Beobachtungen

Unterschiede

- Verwendung der k nächsten Nachbarn vs. Gewichtung der Nachbarn über Kern
- Gleichzeitige Verwendung mehrerer Semimetriken und Kovariablen möglich vs. nur einer Semimetrik auf einer Kovariable möglich
- → Interpretierbares Modell über die Gewichte c_j vs. Bewertung verschiedener Semimetriken nicht in einem Modell möglich

- Logistisches/Multinomiales funktionales Regressionsmodell
- Reduktion der Dimensionalität, z.B. mithilfe funktionaler Hauptkomponentenanalyse (Besse et al.; 1997) oder Diskretisierung der Daten und anschließender Auswertung mit beliebigem multivariaten Klassifikationsverfahren
- Bagnall, A., Lines, J., Bostrom, A., Large, J. and Keogh, E. (2016). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, Data Mining and Knowledge Discovery pp. 1-55 Bagnall et al. (2016)

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten
 - Nonparametric Functional Kernel Estimator
 - Nächste Nachbarn Ensemble
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

Vergleich der Methoden

Tabelle: Mittlere Missklassifikationsrate (MMCR) in % für das Nächste Nachbarn Ensemble (oben, in schwarz) und den Nonparametric Functional Kernel Estimator (unten, in blau) auf den Berkeley Growth Study Daten. Als Anzahl nächster Nachbarn wurden $k = 1, 5, 11$ verwendet, die Bandbreite h wurde über Kreuzvalidierung optimal gewählt.

	Euklid.	Maximum	Minimum	Alle Semimtr.
Original	3.2 8.7	12.3 19.9	43.9 46.6	3.2 -
1. Ableitung	7.5 11.9	55.8 49.8	28.9 33.4	7.5 -
2. Ableitung	5.8 25.1	32.8 34.7	52.6 48.6	5.7 -
Alle Ablt.	6.0 -	14.1 -	30.3 -	5.7 -

Interpretation des Nächste Nachbarn Ensembles

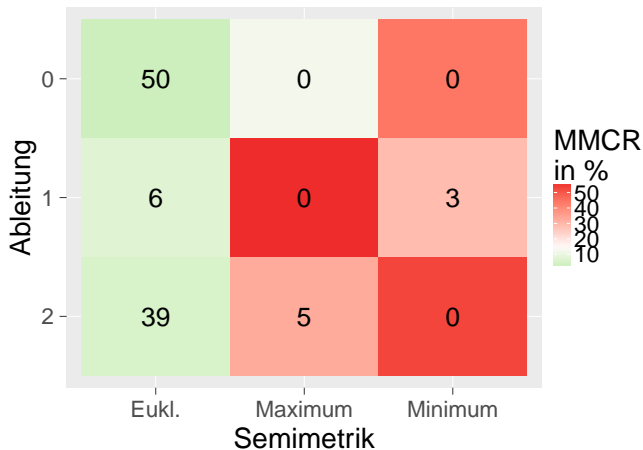


Abbildung: Ensemblegewichte des Nächste Nachbarn Ensembles in %. Die Hintergrundfarbe kodiert die Missklassifikationsrate (MMCR) in %.

- 1 Motivation
- 2 Zentrale Konzepte
 - Funktionale Datenanalyse
 - Klassifikation
 - Semimetriken
- 3 Methoden zur Klassifikation funktionaler Daten
 - Nonparametric Functional Kernel Estimator
 - Nächste Nachbarn Ensemble
 - Kontrastierung und Einordnung der beiden Methoden
- 4 Vergleich der Methoden
- 5 Zusammenfassung & Diskussion

- Nächste Nachbarn Ensemble und Nonparametric Functional Kernel Estimator sind durch die Wahl der Semimetrik(en) sehr flexibel
- Nächste Nachbarn Ensemble ist praktischer in Anwendung wenn optimale Semimetrik unbekannt
- Nächste Nachbarn Ensemble kann mehrere funktionale Kovariablen berücksichtigen

- Dynamic Time Warping als weitere Semimetrik
 - Distanz der gewarpten Funktionen
 - Distanz der Warping Funktionen
- Auf funktionalen Hauptkomponenten basierende Semimetriken
- Flexiblere Wahl der Ensemblegewichte c_l , z.B. über Random Forest

Vielen Dank!



www.matembezi.ch/reisen/suedamerika/patagonien

- Bagnall, A., Lines, J., Bostrom, A., Large, J. and Keogh, E. (2016). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, *Data Mining and Knowledge Discovery* pp. 1–55.
- Besse, P. C., Cardot, H. and Ferraty, F. (1997). Simultaneous non-parametric regressions of unbalanced longitudinal data, *Computational Statistics & Data Analysis* **24**(3): 255–270.
- Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach, *Computational Statistics & Data Analysis* **44**(1): 161–173.
- Fuchs, K., Gertheiss, J. and Tutz, G. (2015). Nearest neighbor ensembles for functional data with interpretable feature selection, *Chemometrics and Intelligent Laboratory Systems* **146**: 186–197.
- Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years., *Publications in child development. University of California, Berkeley* **1**(2): 183.

Zusätzliche Materialien

Auswirkung der Gewichte im Ensemble

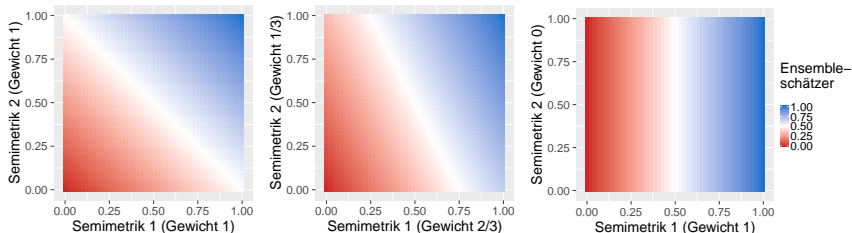


Abbildung: Auswirkung verschiedener Gewichte zweier Semimetriken auf den Ensemble Schätzer. Abgebildet ist die prognostizierte Wahrscheinlichkeit für eine feste Klasse.

Schätzung der Ensemblegewichte über Random Forest

Interaktionen der Semimetriken durch Verwendung eines Random Forest für das Ensemble anstelle der gewichteten Summe

Tabelle: Mittlere Missklassifikationsrate (MMCR) in % für das Nächste Nachbarn Ensemble (oben, in schwarz) und das Random Forest Ensemble (unten, in blau) auf den Berkeley Growth Study Daten. Als Anzahl nächster Nachbarn wurden $k = 1, 5, 11$ verwendet.

	Euklid.	Maximum	Minimum	Alle Semimtr.
Original	3.2	12.3	43.9	3.2
	3.7	13.2	44.1	3.9
1. Ableitung	7.5	55.8	28.9	7.5
	8.5	49.3	32.9	9.0
2. Ableitung	5.8	32.8	52.6	5.7
	6.8	33.5	50.5	7.6
Alle Abt.	6.0	14.1	30.3	5.7
	4.7	13.5	28.9	4.6

Tabelle: Mittlere Missklassifikationsrate (MMCR) in % der Nonparametric Functional Kernel Estimators unter verwendung der Distanz der gewarpten Funktionen und der Warping-Funktionen.

	Gewarpte Funktion	Warping Funktion
Original	15.9	16.6
1. Ableitung	27.6	19.7
2. Ableitung	29.1	20.3