

Abschlussbericht – Kerndichteschätzer

Thomas Fraiss, Maren Heinemann, Tobias Neuschäfer

Beschreibung des Projekts

Ziel des Projekts war die Implementierung eines Kerndichteschätzers sowie dreier Methoden zur automatisierten Bandbreitenwahl. Außerdem wurde für die Generierung eines Zufallsdatensatzes eine Rejection-Sampling-Funktion nach dem Beispiel auf Wikipedia¹ erstellt. Bei den Methoden zur Bandbreitenwahl handelt es sich um die Cross-Validation-Methode und die Goldenshluger-Lepski-Methode aus „Nonparametric Estimation“². Die dritte Methode, die PCO-Methode, wird von Lacour et al. (2017)³ beschrieben, für das Projekt wurde jedoch hauptsächlich der Code zu der entsprechenden Methode von Varet et al. (2018)⁴ verwendet. Teil der Aufgabenstellung war außerdem ein Vergleich der drei beschriebenen Methoden.

Zusammenfassung der Organisation & angewendete Methoden aus VL

Den offiziellen Projektstart haben wir aufgrund der Klausurenphase auf den 24.02. gelegt und auch relativ schnell mit der ersten Aufgabe, der Implementierung der Rejection-Sampling-Funktion begonnen. Während der drei Wochen, die wir für das Projekt zur Verfügung hatten, haben wir uns mehrmals über den aktuellen Stand der Dinge ausgetauscht, ansonsten jedoch eher einzeln gearbeitet. Die Aufgaben wurden je nach Können übernommen. Über das GIT-Repository haben wir unsere Ergebnisse geteilt und an den Lösungen der anderen weitergearbeitet. Bei der Erstellung und Bearbeitung des Pakets haben wir uns an der Vorgehensweise aus der Vorlesung orientiert.

Beiträge einzelner Gruppenmitglieder

Marens Beitrag zum Projekt war hauptsächlich die Vorarbeit, also die Übersetzung des Codes aus dem Buch in R sowie der Einleitungsteil und Schnitt des Videos. „Was gut war, ist dass die Aufgabenstellung klar definiert war und durch den Code in den Büchern viel Hilfestellung gegeben war. Trotzdem finde

¹ https://en.wikipedia.org/wiki/Rejection_sampling

² Comte, F.: *Nonparametric Estimation*. Spartacus-Idh (2017)

³ Lacour et al, Estimator selection: a new method with applications to kernel density estimation, (2017), <https://arxiv.org/abs/1607.05091>

⁴ Varet et al, Numerical performance of Penalized Comparison to Overfitting for multivariate kernel density estimation, (2018), <https://arxiv.org/abs/1902.01075>

ich aufgrund des Gesamtaufwands während des Semesters und der Projektphase, sollte der Kurs mehr als 8 ECTS geben.“

Thomas hat aus dieser Vorarbeit die fertigen Funktionen gebaut. Außerdem hat er die Roxygen2-Kommentare geschrieben, die erste Version der Vignette erstellt und einen Wissensbeitrag zum Video geliefert. „Was gut geklappt hat, war die Funktionen aus dem Buch zu übernehmen. Frustrierend war, dass die Code-Beispiele an expliziten Kernen, anstelle von allgemeinen Kernen gemacht wurde, was dazu führte, dass man einige Stellen im Code umschreiben musste, zum Beispiel die Konvolution von zwei Kernen in der Goldenshluger-Lepski-Methode.“

Tobias hat den Kerndichteschätzer geschrieben sowie das Paket, inklusive Tests, erstellt und sich um den Feinschliff gekümmert. Dazu gehörte auch die Funktionalität vorgefertigte Kerne nutzen zu können, die Anfertigung des Grundgerüsts der Roxygen2-Kommentare und die Verbesserung der Vignette. Außerdem hat er das Tutorial zur Benutzung des Pakets und den Vergleich der drei Methoden zur Bandbreitenwahl in der Abschlusspräsentation erstellt.