

Statistische Verfahren SS 2019

Projekt 4 – Nahinfrarotspektroskopie I

Problemstellung:

Herleitung von Kalibrierungsfunktionen zur Schätzung des Bodenkohlenstoffgehalts mit Hilfe der Nahinfrarotspektroskopie

Datensatz: (NIR.csv, Dr. A. Don, Heinrich von Thünen-Institut Braunschweig)

Der Datensatz enthält Daten zum Bodennitrostoffgehalt (N) sowie die dazugehörigen Nahinfrarotspektren, d.h. die gemessenen Reflexionswerte bei verschiedenen Wellenlängen aus dem Nahinfrarotbereich, für 534 Bodenproben. Diese Daten sollen zur Herleitung von Prognosefunktionen verwendet werden, die die Schätzung des nur aufwändig bestimmbareren Bodenparameters N aus den viel leichter messbaren Nahinfrarotspektren gestattet.

Aufgaben:

- Bestimmen Sie mit Hilfe eines geeigneten Modellselektionsverfahrens Vorhersagemodelle für den Bodennitrostoffgehalt N.

Simulationsaufgabe:

- Wählen Sie eines der im ersten Teil hergeleiteten Prognosemodelle als Grundlage zur Simulation von Pseudobeobachtungen. Entscheiden Sie sich für jeden simulierten Pseudo-Datensatz basierend auf Mallows C_p -Kriterium für ein bestes Modell. Vergleichen Sie den aus dem minimalen C_p -Wert geschätzten Vorhersagefehler mit dem theoretisch erwarteten Vorhersagefehler SPSE des zur Simulation verwendeten wahren Modells. Untersuchen Sie dabei den Einfluss des Stichprobenumfangs (indem Sie aus dem Originaldatensatz zufällig Spektren für die Designmatrix der Simulation auswählen).