h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES
fbmn
FACHBEREICH MATHEMATIK
UND NATURWISSENSCHAFTEN

Antje Jahn
Nichtparametrische und nichtlineare Modelle
Sommersemester 2019

Arbeitsblatt 3

A 1

Angenommen, die Zufallsvariablen

$$X \sim U[1,5] \tag{1}$$

und Y folgen dem statistischen (datengenerierenden) Modell

$$Y = f(x) + \epsilon = \beta \cdot log(X) + \epsilon \text{ mit } \epsilon \sim N(0,1) \tag{2}$$

Sie möchten die Frage beantworten, wie groß jeweils der erwartete quadratische Vorhersagefehler aus einem (missspezifizierten) linearen Analysemodell

$$Y = \beta_0^L + \beta_1^L \cdot X + \epsilon$$

und einer polynomialen Regression vom Grad 5

$$Y = \beta_0^P + \beta_1^P \cdot X + \beta_2^P \cdot X^2 + \cdots + \beta_5^P \cdot X^5 + \epsilon$$

an der Stelle $x = E(X) = 3$ ist. Setzen Sie den Parameter im datengenerierenden Modell $\beta = 4$ und führen Sie eine Simulationsstudie durch:

a) Simulieren Sie einen Datensatz mit $n = 100$ Beobachtungen und 2 Variablen X und Y entsprechend Modell (2).
b) Führen Sie eine lineare Regression und eine polynomiale Regression vom Grad 5 durch und berechnen Sie für jedes Regressionsmodell $\hat{f}(x)$.
c) Simulieren Sie eine weitere Zufallszahl $y \sim f(X) + \epsilon$ für ein festes $x = 3$ und berechnen Sie den quadratischen Vorhersagefehler $(y - \hat{f}(x))^2$ für jedes der beiden Modelle.
d) Wiederholen Sie die Schritte b)-d) 10000 Mal und berechnen Sie für jedes Regressionsmodell den Erwartungswertschätzer für den Bias, $E(\hat{f}(x)) - f(x)$, den Varianzschätzer für $Var(\hat{f}(x))$ und den Erwartungswertschätzer für den erwarteten quadratischen Vorhersagefehler $E[(Y - \hat{f}(x))^2]$.
e) Betrachten Sie Ihre Schätzer für Bias, Varianz und quadratischen Vorhersagefehler:
   i) Welches der beiden Modelle zeigt den kleineren Bias, welches die kleinere Varianz und welches den kleineren quadratischen Vorhersagefehler?

ii) Welchen Zusammenhang sollten diese drei Größen für jedes der beiden Regressionsmodelle gleichermaßen aufzeigen? Überprüfen Sie diesen Zusammenhang in Ihren Simulationsergebnissen.

iii) In welche Richtung würden sich Bias und Varianz voraussichtlich ändern, wenn die Fallzahl statt $n = 100$ nur noch $n = 20$ beträgt? Überprüfen Sie Ihre Vermutung anhand einer zweiten Simulationsstudie. Welches Modell zeigt nun den kleineren quadratischen Vorhersagefehler?

f) Erklären Sie die Ergebnisse aus e).