Übungsaufgaben

16) Diskutieren Sie qualitativ wie sich bei der Kreuzvalidierungsschätzung $\widehat{R}_{KV}(\widehat{y})$ des erwarteten Risikos $R(\widehat{y})$ die Anzahl der Falten $k = 5$, $k = 10$ und $k = N$ (Leave-One-Out Kreuzvalidierung) auf den Bias und die Varianz von $\widehat{R}_{KV}(\widehat{y})$ auswirkt. Gehen Sie dabei auf die folgenden Punkte ein:

   a) Größe der Trainingsdatensätze in Abhängigkeit der Anzahl k von Falten.
   b) Verzerrung (Bias) des Kreuzvalidierungsschätzers $\widehat{R}_{KV}(\widehat{y})$.
   c) Schwankung (Varianz) des Kreuzvalidierungsschätzers $\widehat{R}_{KV}(\widehat{y})$.
   d) Praktische Empfehlungen für die Wahl von k.

Übungsaufgaben

17) Gegeben seien die (geordnete) Stichprobe

$$12, 15, 18, 21, 22, 25, 30, 32. \tag{2}$$

Zu schätzen sei das 25%-Quantil $\theta = q_{0,25}$ der zugrundeliegenden Verteilung F.

a) Bestimmen Sie mit (2) die Stichprobenschätzung $\widehat{\theta}$ für das 25%-Quantil $\theta = q_{0,25}$ von F.

<u>Hinweis:</u> Als Schätzung eines p-Quantils einer Stichprobe $x_1, x_2, \ldots, x_N$ ist das entsprechende empirische Quantil $x_{(p)}$ definiert durch: für wenigstens pN der Beobachtungen gilt $x_n \leq x(p)$ und für wenigstens $(1-p)N$ der Werte gilt $x_n \geq x(p)$.

b) Es wurden die folgenden unabhängigen $B = 5$ Bootstrap-Stichproben mit Zurücklegen vom Umfang $N = 8$ erzeugt:

$$\begin{aligned}
\mathbf{x}_1^* &= (12, 12, 15, 18, 21, 25, 30, 32)\\
\mathbf{x}_2^* &= (18, 18, 21, 22, 22, 25, 32, 32)\\
\mathbf{x}_3^* &= (12, 15, 15, 18, 21, 21, 22, 25)\\
\mathbf{x}_4^* &= (15, 18, 21, 22, 25, 25, 30, 30)\\
\mathbf{x}_5^* &= (18, 21, 21, 22, 25, 30, 30, 32)
\end{aligned}$$

Berechnen Sie für jede Bootstrap-Stichprobe die Realisierung $\widehat{\theta}_b^*$.

Übungsaufgaben

c) Bestimmen Sie die Bootstrap-Schätzung für das 25%-Quantil $\theta = q_{0,25}$ und damit die Bootstrap-Schätzung für den Bias von $\widehat{\theta}$.

d) Bestimmen Sie die Bootstrap-Schätzung für die Varianz und die Standardabweichung von $\widehat{\theta}$.

e) Konstruieren Sie ein 95%-Bootstrap-Konfidenzintervall für θ, indem Sie die 2,5%- und 97,5%-Quantile der (simulierten) empirischen Verteilung $\left\{\widehat{\theta}_1^*, \widehat{\theta}_2^*, \widehat{\theta}_3^*, \widehat{\theta}_4^*, \widehat{\theta}_5^*\right\}$ als Intervallgrenzen verwenden.

   Hinweis: Bei $B = 5$ Realisierungen $\widehat{\theta}_b^*$ dienen die kleinste und die größte Realisierung als Approximation.

f) Diskutieren Sie warum Bootstrapping besonders für Quantilschätzungen nützlich ist und weshalb asymptotische Normalapproximationen hier oft versagen?

18) Eine Versicherungsgesellschaft betrachtet die Schadenhöhen (in Tsd. CHF) einer bestimmten Sparte. Es liegt die folgende (geordnete) Stichprobe vor

$$8, 10, 15, 20, 25, 30, 60. \tag{3}$$

Zu schätzen sei die Varianz σ^2 der zugrundeliegenden Verteilung F.

a) Berechnen Sie den Stichprobenmittelwert $\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$ und die korrigierte Stichprobenvarianz

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \overline{x})^2.$$

b) Es wurden die folgenden unabhängigen $B = 5$ Bootstrap-Stichproben mit Zurücklegen vom Umfang $N = 7$ erzeugt:

$$\begin{aligned}
\mathbf{x}_1^* &= (8, 10, 15, 20, 25, 30, 60)\\
\mathbf{x}_2^* &= (8, 8, 10, 15, 20, 25, 30)\\
\mathbf{x}_3^* &= (25, 25, 30, 30, 20, 15, 10)\\
\mathbf{x}_4^* &= (60, 60, 60, 30, 25, 20, 15)\\
\mathbf{x}_5^* &= (8, 10, 15, 20, 25, 30, 60)
\end{aligned}$$

Berechnen Sie für jede Bootstrap-Stichprobe die Realisierung $\widehat{\sigma^2}_b^*$.

Übungsaufgaben

c) Bestimmen Sie die Bootstrap-Schätzung für die Varianz σ^2 und damit die Bootstrap-Schätzung für den Bias von $\widehat{\sigma^2}$.

d) Bestimmen Sie die Bootstrap-Schätzung für die Varianz und die Standardabweichung von $\widehat{\sigma^2}$.

e) Diskutieren Sie, welche Besonderheiten bei der Bootstrap-Approximation der Verteilung von $\widehat{\sigma^2}$ auftreten, wenn die Ausgangsstichprobe einen starken Ausreißer (hier $x_7 = 60$) enthält.