UNI FREIBURG

Maschinelles Lernen aus Stochastischer Sicht — Sommersemester 2024

Vorlesung: Prof. Dr. Thorsten Schmidt
Übung: Moritz Ritter

Übungsblatt 5

Abgabe: Freitag, 31.05.2024, 18:00 Uhr.

Aufgabe 1 (4 Punkte)**.** Verwenden Sie die Ergebnisse aus Aufgabe 4 Blatt 4, um zu zeigen, dass

$$B_{\sqrt{d}\hat{s}t_{n-d,1-\frac{\alpha}{2d}}}(\hat{\beta}) := \{x \in \mathbb{R}^d | \|x-\hat{\beta}\| < \sqrt{d}\hat{s}t_{n-d,1-\frac{\alpha}{2d}}\}$$

ein Konfidenzbereich für β zum Parameter α ist. Dabei bezeichnet $t_{n-d,1-\frac{\alpha}{2d}}$ das das $\frac{\alpha}{2d}$-Quantil der t_{n-d} Verteilung.

Aufgabe 2 (4 Punkte)**.** Visualisieren Sie die Konfidenzbereiche, sowie die Regressionsgerade in einem Beispiel ihrer Wahl. Wieso unteschieden sich diese Konfidenzbereiche? Wieso liegen viele der Datenpunkte nicht in den Konfidenzbereichen?

Aufgabe 3 (4 Punkte)**.** Sei $G\colon \mathbb{R}^d \to \mathscr{G}$ mit $\mathscr{G} = \{1,\dots,K\}$ eine Klassifizierung eines Vektors $x \in \mathbb{R}^d$ in die Klasse $k \in \mathscr{G}$. Weiter sei $Y = (Y_1,\dots,Y_K)$ mit $Y_k = 1$ und 0 sonst, falls die Klassifizierung der Kovariablen $G(X_1,\dots,X_d) = k$ der k'ten Gruppe entspricht. Eine Klassifizierung von $x \in \mathbb{R}^d$ basierend auf den Daten kann mittels linearer Regression wie folgt durchgeführt werden:

$$\hat{G}(x) := \operatorname{argmin}_k \|\hat{f}(x) - e_k\|, \tag{1}$$

für $\hat{f} = (\hat{f}^1,\dots,\hat{f}^K)$ mit $\hat{f}^k(x) = \hat{\beta}_0^k + \langle\hat{\beta}^k, x\rangle$, wobei $\hat{\beta}^k$ den KQS für das lineare Modell $Y^k = \langle\beta, X\rangle + \varepsilon$ bezeichnet und e_k den k'ten Einheitsvektor in $\mathbb{R}^K$.

1. Erzeugen Sie Daten $X^k \sim \mathcal{N}(k,1)$ für $k \in \mathscr{G} = \{1,2\}$ und implementieren Sie den Klassifizierer $\hat{G}$, welcher durch (1) definiert wird. Plotten Sie die Funktionen $\hat{f}_k$ und die daraus resultierenden Klassifizierungsbereiche.
2. Erzeugen Sie Daten $X^k \sim \mathcal{N}(k,1)$ für $k \in \mathscr{G} = \{1,2,3\}$ und implementieren Sie den Klassifizierer $\hat{G}$, welcher durch (1) definiert wird. Plotten Sie die Funktionen $\hat{f}_k$ und die daraus resultierenden Klassifizierungsbereiche. Welche Probleme sehen Sie durch diese Art von Klassifizierung?
3. Transformieren Sie die Kovariablen, wie zu vor von linearer zu polynomialer Regression, um auch für $|\mathscr{G}| \geq 3$ eine gute Schätzung zu erhalten und implementieren Sie diese für 2..

Hinweis zu Aufgabe 2 und 3: Ihre Abgabe soll sowohl aus dem Code (2 Punkte) als auch einer übersichtlichen Darstellung durch Plots und einer Erklärung der Resultate (2 Punkte) bestehen.