

Machine Learning

Übungsblatt 2

25 Punkte

Aufgabe 1. *Beispiel: Würfel – Bedingte Wahrscheinlichkeit*

7 P.

Gegeben sei ein fairer 6-seitiger Würfel der zweimal geworfen wird. Es bezeichnen X_1 und X_2 die jeweils gewürfelte Augenzahl und $S = X_1 + X_2$ deren Summe.

- (a) Sind die folgenden Paare an Zufallsvariablen abhängig oder unabhängig: (i) X_1 und X_2 , (ii) X_1 und S ?
 (b) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}[S = 8]$.

Bestimmen Sie bedingte Wahrscheinlichkeit $\mathbb{P}[X_1 = 2|S = 8]$ auf zwei verschiedene Arten.

- (b) Direkt, d.h., durch Betrachten der möglichen Werte von X_1 , sodass $S = 8$.
 (c) Indirekt, durch Anwendung des Satzes von Bayes.

Aufgabe 2. *Mehrdimensionale Normalverteilung – Teil I*

6 P.

Die d -dimensionale standard Normalverteilung $\mathcal{N}(\mathbf{0}, \mathbf{1})$ ist die Wahrscheinlichkeitsverteilung für Zufallsvektoren $\mathbf{X} = (X_1, \dots, X_d)$, wobei alle Komponenten voneinander unabhängig sind und einer eindimensionalen standard Normalverteilung folgen, d.h., $X_i \sim \mathcal{N}(0, 1)$.

Ähnlich zum eindimensionalen Fall schreiben wir $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ falls $(Y_1, \dots, Y_d) = \mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mu}$ mit $\mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ und $(X_1, \dots, X_k) = \mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$. Hierbei ist $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ ein Vektor und $\mathbf{A} \in \mathbb{R}^{d \times k}$ eine Matrix mit Einträgen $a_{ij} \in \mathbb{R}$.

- (a) Zeigen Sie zunächst, dass für beliebige Zufallsvariablen $\mathbf{Z} \in \mathbb{R}^k$, Matrizen $\mathbf{M} \in \mathbb{R}^{d \times k}$ und Vektoren $\mathbf{v} \in \mathbb{R}^d$ gilt dass

$$\mathbb{E}[\mathbf{M}\mathbf{Z} + \mathbf{v}] = \mathbf{M}\mathbb{E}[\mathbf{Z}] + \mathbf{v} .$$

Hinweis: Betrachten Sie die i -te Komponente des Zufallsvektors $\mathbf{M}\mathbf{Z} + \mathbf{v}$ und nutzen Sie Aufgabe 3(a) von Blatt 1.

- (b) Folgern Sie, dass $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}]$ und dass $\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top]$.

Aufgabe 3. *Parameterschätzung*

12 P.

Gegeben seien unabhängige Zufallsvariablen X_1, \dots, X_n mit jeweils gleicher Dichtefunktion

$$f_\theta : [0, 1] \rightarrow \mathbb{R}^+, x \mapsto \theta x^{\theta-1} .$$

Hierbei ist $\theta \geq 1$ eine beliebige Zahl, die die Wahrscheinlichkeitsdichte bestimmt.

- (a) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}[X_1 \in (0, a) \text{ und } X_2 \in (0, b) \text{ und } X_3 \in (0, c)]$.
 (b) Die gemeinsame Dichte von X_1, \dots, X_n ist die Funktion $g_\theta : [0, 1]^n \rightarrow \mathbb{R}^+$, so dass

$$\mathbb{P}[(X_1, \dots, X_n) \in (a_1, b_1) \times \dots \times (a_n, b_n)] = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} g_\theta(x_1, \dots, x_n) dx_1 \dots dx_n . \quad (1)$$

Bestimmen Sie g_θ .

Wir beobachten eine Stichprobe $x_1 = 0.8, x_2 = 0.85, x_3 = 0.9$.

- (c) Zeichnen Sie die Werte x_i und die Funktionsgraphen der Dichtefunktionen f_2 und f_4 in ein Koordinatensystem ein.
 (d) Welcher der beiden Dichtefunktionen ist besser geeignet um die Beobachtungen zu modellieren. Begründen Sie ihre Antwort. Ihre Begründung sollte *keine Rechnung* enthalten.

Wir suchen nun den Wert von θ , unter dem die Beobachtungen am Wahrscheinlichsten sind, d.h., das Maximum der Funktion

$$L : [1, \infty) \rightarrow \mathbb{R}^+, \theta \mapsto g_\theta(x_1, \dots, x_n)$$

Hierbei sind x_1, \dots, x_n bekannt, aber beliebig.

- (e) Begründen Sie, warum es ausreicht das Maximum der Funktion $\theta \mapsto \log(L(\theta))$ zu finden und warum dies einfacher ist.
- (f) Berechnen Sie das globale Maximum $\hat{\theta}$ von $\log L$.
Hinweis: Berechnen Sie hierfür das lokale Maximum von $\log L$, und begründen Sie, warum es auch das globale Maximum ist.
- (g) Geben Sie $\hat{\theta}$ für den Fall $x_1 = 0.8, x_2 = 0.85, x_3 = 0.9$ an.