Universität Salzburg | Florian Graf

Machine Learning

Übungsblatt **4** | 20 Punkte

Aufgabe 1. *MLE – Univariate Normalverteilung* 12 P.

Gegeben seien Beobachtungen $x_1, \dots, x_n, \in \mathbb{R}$. Wir nehmen an, dass die Beobachtungen von einer normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ stammen.

(a) Schätzen Sie die Parameter μ und σ^2 gemeinsam mittels der Maximum-Likelihood Methode. Überprüfen Sie, dass tatsächlich ein globales Maximum vorliegt.

Hinweis: In diesem Fall ist die log-Likelihoodfunktion eine Funktion $\log L : \mathbb{R}^2 \to \mathbb{R}$. Eine notwendige Bedingung an die lokalen Maxima $(\hat{\mu}, \hat{\sigma}^2)$ einer solchen Funktion ist, dass der Gradient

$$\nabla \log L(\mu, \sigma^2) := \begin{pmatrix} \frac{\partial \log L(\mu,\sigma^2)}{\partial \mu} \\ \frac{\partial \log L(\mu,\sigma^2)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{für } \mu = \hat{\mu} \text{ und } \sigma^2 = \hat{\sigma}^2 \ ,$$

und dass die Hessematrix

$$H_{\log L}(\mu, \sigma^2) := \begin{pmatrix} \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \mu \partial \mu} & \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix} \quad \text{an der Stelle } (\mu, \sigma^2) = (\hat{\mu}, \hat{\sigma}^2) \text{ negativ definit ist.}$$

Da (üblicherweise) unsere Beobachtungen $x_1, \dots, x_n$ zufallsbehaftet sind, sind es auch die Parameterschätzer $\hat{\theta}$ und wir können sie als Zufallsvariablen auffassen.

- Ein Schätzer $\hat{\theta}$ für den Parameter θ heißt *verzerrt*, falls $\mathbb{E}[\hat{\theta}] \neq \theta$ gilt.
- Ein Schätzer heißt *erwartungstreu* oder *unverzerrt*, falls $\mathbb{E}[\hat{\theta}] = \theta$.
- Ein Schätzer heißt *konsistent*, falls er mit wachsendem Stichprobenumfang n dem wahren Parameterwert θ (mit hoher Wahrscheinlichkeit) immer näher kommt.

(b) Zeigen Sie, dass der Maximum-Likelihood Schätzer $\hat{\mu}$ für μ erwartungstreu ist.

(c) Ist $\hat{\mu}$ konsistent? Betrachten Sie dazu $\mathrm{Var}(\hat{\mu})$.

(d) Zeigen Sie, dass der Maximum-Likelihood Schätzer $\hat{\sigma}^2$ für σ^2 verzerrt ist und geben Sie einen unverzerrten Schätzer für σ^2 an.

Aufgabe 2. *Gradienten* 4 P.

Berechnen Sie den Gradienten und die Hessematrix der folgenden Funktionen $\mathbb{R}^d \to \mathbb{R}$. Berechnen Sie zunächst die partiellen Ableitungen und schreiben Sie anschließend den Gradienten (bzw. die Hessematrix) als Matrix-Vektor-Produkt.

(a) $f_1(x) = \mathbf{w}^\top \mathbf{x}$ mit $\mathbf{w}, \mathbf{x} \in \mathbb{R}^d$.

(b) $f_2(x) = \frac{1}{2}\|\mathbf{x}\|^2$ mit $\mathbf{x} \in \mathbb{R}^d$.

(c) $f_3(x) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x}$ mit $\mathbf{x} \in \mathbb{R}^d$ und $A \in \mathbb{R}^{d \times d}$.

(d) $f_4(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ mit $\mathbf{x}, \mathbf{b} \in \mathbb{R}^d$ und $A \in \mathbb{R}^{d \times d}$.

Aufgabe 3. *MLE - Mehrdimensionale Normalverteilung* 4 P.

Gegeben seien n Beobachtungen $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$. Wir nehmen an, dass die Beobachtungen von einer Normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \Sigma)$ stammen, wobei $\boldsymbol{\Sigma}$ bekannt ist und $\ker(\boldsymbol{\Sigma}) = \{\mathbf{0}\}$.

(a) Zeigen Sie, dass der Maximum-Likelihood Schätzer $\hat{\boldsymbol{\mu}}$ für $\boldsymbol{\mu}$ die Größe $\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$ minimiert und interpretieren Sie diesen Zusammenhang.

(b) Bestimmen Sie $\hat{\boldsymbol{\mu}}$. (Wie immer ist es Teil der Aufgabe zu überprüfen, dass die Likelihoodfunktion an der Stelle $\hat{\boldsymbol{\mu}}$ tatsächlich ihr Maximum annimmt.)