

**Machine Learning**

## Übungsblatt 6

24 Punkte

**Aufgabe 1.** MAP Schätzer – Gauß Prior

12 P.

Es sei  $\mathcal{D} = \{x_1, \dots, x_n\} \in \mathbb{R}$  eine Stichprobe einer univariaten Normalverteilung mit bekannter Varianz  $\sigma^2$  und unbekanntem Mittelwert  $\mu$ . Außerdem nehmen wir eine A-priori-Verteilung für  $\mu \sim \mathcal{N}(m, s^2)$  an, wobei  $m \in \mathbb{R}$  und  $s^2 > 0$  bekannt und fest sind.

- (a) Zeigen Sie, dass der Logarithmus der A-posteriori-Verteilung durch

$$\log p(\mu|\mathcal{D}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{1}{2s^2} (\mu - m)^2 + \text{const}$$

gegeben ist, wobei const einen Term beinhaltet, der nicht von  $\mu$  abhängt.

- (b) Zeigen Sie, dass

$$\hat{\mu}_{\text{MAP}} = \frac{ns^2}{ns^2 + \sigma^2} \frac{1}{n} \sum_{i=1}^n x_i + \frac{\sigma^2}{ns^2 + \sigma^2} m$$

ein Maximum (der Dichte) der A-posteriori Verteilung ist.

- (c) Zeigen Sie, dass für  $m = 0$  die Ungleichung  $|\hat{\mu}_{\text{MAP}}| \leq |\hat{\mu}_{\text{MLE}}|$  gilt.
- (d) Zeigen Sie dass für beliebiges  $m \in \mathbb{R}$  gilt, dass  $\hat{\mu}_{\text{MAP}} = \hat{\mu}_{\text{MLE}}$  falls  $\hat{\mu}_{\text{MLE}} = m$ .  
Zeigen Sie auch, dass für  $\hat{\mu}_{\text{MLE}} < m$  gilt dass  $\hat{\mu}_{\text{MAP}} \in (\hat{\mu}_{\text{MLE}}, m)$ , bzw.  $\hat{\mu}_{\text{MAP}} \in (m, \hat{\mu}_{\text{MLE}})$  falls  $\hat{\mu}_{\text{MLE}} > m$ . (Es genügt, wenn Sie einen Fall aufschreiben.)
- (e) Zeigen Sie, dass mit zunehmendem Stichprobenumfang  $n$ , der MAP-Schätzer für  $\mu$  gegen dessen MLE-Schätzer  $\hat{\mu}_{\text{MLE}}$  konvergiert.
- (f) Wogegen konvergiert der MAP-Schätzer mit zunehmender a priori Varianz  $s^2$  (für festes  $n$ ).
- (g) Wogegen konvergiert der MAP-Schätzer mit abnehmender a priori Varianz  $s^2$  (für festes  $n$ ).
- (h) Interpretieren Sie die Ergebnisse von (d) bis (h). Fassen Sie sich kurz (jeweils 2 bis 3 Sätze).

**Aufgabe 2.** Laplace Verteilung

4 P.

Die Laplaceverteilung mit Parametern  $\mu \in \mathbb{R}$  und  $\sigma > 0$  ist definiert über die Dichte

$$p(x|\mu, \sigma) = \text{Laplace}(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right).$$

- (a) Zeigen Sie zunächst, dass  $\mathbb{E}[X] = 0$  wenn  $X \sim \text{Laplace}(0, \sigma)$ . Folgern Sie daraus, dass  $\mathbb{E}[X] = \mu$  wenn  $X \sim \text{Laplace}(x|\mu, \sigma)$ . Sie dürfen dazu ohne Beweis verwenden, dass  $\Pr[X \in \mathbb{R}] = 1$ .
- (b) Skizzieren Sie die Funktionsgraphen der Wahrscheinlichkeitsdichten von  $X \sim \text{Laplace}(0, 1)$  und  $Y \sim \mathcal{N}(0, 1)$  in einer handschriftlichen Zeichnung. Achten Sie dabei besonders auf das Verhalten der Dichten in der Nähe von 0.

Tipp: Visualisieren Sie zunächst die Dichten mit einer Software (z.B. WolframAlpha).

**Aufgabe 3.** MAP Schätzer – Laplace Prior

8 P.

Wir sind erneut in dem Setting von Aufgabe 1 (d.h. Stichprobe  $\mathcal{D} = \{x_1, \dots, x_n\} \in \mathbb{R}$  sei eine Stichprobe einer univariaten Normalverteilung mit bekannter Varianz  $\sigma^2$  und unbekanntem Mittelwert  $\mu$ ). Allerdings nehmen wir nun einen Laplace Prior  $\mu \sim \text{Laplace}(0, s)$  an, wobei  $s > 0$  bekannt und fest ist.

- (a) Zeigen Sie, dass der Logarithmus der A-posteriori-Verteilung durch

$$\log p(\mu|\mathcal{D}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{1}{2s} |\mu| + \text{const}$$

gegeben ist, wobei const einen Term beinhaltet, der nicht von  $\mu$  abhängt.

- (b) Es sei  $\hat{\mu}_{\text{MLE}}$  der MLE Schätzer von  $\mu$ . Zeigen Sie, dass

$$\frac{d \log p(\mu|\mathcal{D})}{d\mu} = \begin{cases} \frac{n}{\sigma^2} (\hat{\mu}_{\text{MLE}} - \mu) + \frac{1}{2s} & \mu < 0 \\ \frac{n}{\sigma^2} (\hat{\mu}_{\text{MLE}} - \mu) - \frac{1}{2s} & \mu > 0 \end{cases}$$

- (c) Es sei  $\epsilon = \frac{\sigma^2}{2ns}$ . Folgern Sie aus (b), dass

$$\hat{\mu}_{\text{MAP}} = \begin{cases} \hat{\mu}_{\text{MLE}} - \epsilon \text{sign}(\hat{\mu}_{\text{MLE}}), & \text{falls } |\hat{\mu}_{\text{MLE}}| > \epsilon \\ 0 & \text{sonst} \end{cases}.$$

Sie dürfen dafür ohne Beweis annehmen, dass die Funktion  $\mu \mapsto \log p(\mu|\mathcal{D})$  ein Maximum besitzt.

- (d) Interpretieren Sie das Resultat aus (c).