Machine Learning

Aufgabe 1. *MAP Schätzer – Gauß Prior* 11 P.

Es sei $\mathcal{D} = \{x_1, \ldots, x_n\} \in \mathbb{R}$ eine Stichprobe einer univariaten Normalverteilung mit bekannter Varianz σ^2 und unbekanntem Mittelwert μ. 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 μ abhängt.

(b) Bestimmen Sie den Maximierer $\hat{\mu}_{\text{MLE}} := \arg\max_\mu \log p(\mu|\mathcal{D})$ der Dichte der A-Posteriori-Verteilung und schreiben Sie diesen in Abhängigkeit des Maximum-Likelihood Schätzers $\hat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ von μ.

(c) Folgern Sie, dass

$$|\hat{\mu}_{\text{MAP}} - m| \leq |\hat{\mu}_{\text{MLE}} - m| \ ,$$

und dass Gleichheit gilt, falls $m = \hat{\mu}_{\text{MLE}}$.

(d) Folgern Sie auch, dass mit zunehmendem Stichprobenumfang n, der MAP-Schätzer $\hat{\mu}_{\text{MAP}}$ gegen den MLE-Schätzer $\hat{\mu}_{\text{MLE}}$ von μ konvergiert.

(e) Wogegen konvergiert der MAP-Schätzer mit zunehmender a priori Varianz s^2 (für festes n).

(f) Wogegen konvergiert der MAP-Schätzer mit abnehmender a priori Varianz s^2 (für festes n).

(g) Interpretieren Sie die Ergebnisse von (c) bis (h).

Aufgabe 2. *Regularisierung* 4 P.

Wir betrachten ein diskriminatives Klassifizierungsmodell $p(y|\mathbf{x}, \boldsymbol{\theta})$ mit Parametervektor $\boldsymbol{\theta} \in \mathbb{R}^d$.

(a) Zeigen Sie, dass die Maximum-A-Posteriori-Schätzung (MAP) des Parameters $\boldsymbol{\theta}$ unter der Annahme einer isotropen Normalverteilung $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda}\mathbf{I})$ dem Minimum der regularisierten Fehlerfunktion

$$\mathcal{L}(\boldsymbol{\theta}) = \text{NLL}(\boldsymbol{\theta}) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|^2$$

entspricht.

(b) Wir machen nun stattdessen die Annahme, dass die Komponenten θ_i des Vektors $\boldsymbol{\theta}$ unabhängig voneinander Laplaceverteilt sind, also $\theta_i \sim \text{Lap}(0, \frac{1}{\lambda})$. Welcher regularisierten Fehlerfunktion entspricht die Maximum-A-Posteriori-Schätzung des Parameters $\boldsymbol{\theta}$ in diesem Fall?

Hinweis: Die Dichtefunktion einer Laplace- $\text{Lap}(\mu, b)$ mit Parametern μ, b ist gegeben durch

$$y \mapsto \frac{1}{2b}\exp(-\frac{|y - \mu|}{b}) \ .$$