STRUKTURELLE MODELLE IN DER BILDVERARBEITUNG
4. ÜBUNG – LERNEN MARKOVSCHER KETTEN

Aufgabe 1. Ein Objekt kann sich in zwei Zuständen $k = 1,2$ befinden. Die a-priori Wahrscheinlichkeiten $p(k=1)$ und $p(k=2)$ seien bekannt. Die bedingten Wahrscheinlichkeiten für die Merkmale $x \in \mathbb{R}^n$ sind Gaußsch verteilt:

$$p(x|k) = \frac{1}{(\sqrt{2\pi}\sigma_k)^n} \exp\Big[-\frac{||x-\mu||^2}{2\sigma_k^2}\Big].$$

Beide Verteilungen haben dasselbe Zentrum μ aber unterschiedliche Streuungen σ_k. Gegeben sei eine klassifizierte Stichprobe $L = \big((x^1,k^1),(x^2,k^2)\ldots(x^l,k^l)\big)$. Seien die Streuungen σ_k bekannt. Man schätze μ mit Hilfe des Maximum-Likelihood Prinzips.

Hinweis: Betrachten Sie das Beispiel des Lernens von μ aus der Vorlesung. Verfahren Sie hier zunächst analog. Berücksichtigen Sie allerdings, dass sich die σ_k in der jetzigen Aufgabe nicht mehr kürzen lassen.

Aufgabe 2. Die Wahrscheinlichkeitsverteilung einer skalaren Größe $x \in \mathbb{R}$ ist

$$p(x) = C \cdot \exp\big[-\tau|x-\mu|\big]$$

mit reellen Parametern τ und μ.

a) Wie ergibt sich die Normierungskonstante C aus den bekannten Parametern τ und μ?
Hinweis: Nutzen Sie $\int\limits_{-\infty}^{\infty} p(x)dx = 1$.
b) Die Parameter τ und μ sollen nach dem Maximum-Likelihood Prinzip anhand einer Lernstichprobe $L = (x^1,x^2\ldots x^l)$ gelernt werden. Wie ergeben sich daraus die gesuchten Größen?
c) Lösen Sie diese Aufgabe für die Wahrscheinlichkeitsverteilung (mit $\tau > 0$)

$$p(x) = \begin{cases} C \cdot \exp\big[-\tau(x-\mu)\big] & \text{wenn } x \geq \mu, \\ 0 & \text{sonst.} \end{cases}$$

Hinweis: Berücksichtigen Sie, dass bei manchen Werten von μ die Wahrscheinlichkeit bestimmter Muster x^l in der Lernstichprobe $p(x^l) = 0$ ist. Somit ist die Wahrscheinlichkeit der ganzen Lernstichprobe auch $P(L) = 0$ (und folglich nicht maximal). Offensichtlich müssen solche Werte von μ ausgeschlossen werden.

Aufgabe 3. Man betrachte das Markovsche Modell für Zustandsfolgen $y=(y_1,y_2\dots y_n)$, mit $y_i\in K$, bei dem die Parameter nicht unbedingt Wahrscheinlichkeitswerten entsprechen, d.h.

$$p(y)=\frac{1}{Z}\cdot\prod_{i=2}^{n}g_i(y_{i-1},y_i)\cdot\prod_{i=1}^{n}q_i(y_i),$$

mit der Partition Funktion

$$Z=\sum_y\left[\prod_{i=2}^{n}g_i(y_{i-1},y_i)\cdot\prod_{i=1}^{n}q_i(y_i)\right].$$

Für die weiteren Betrachtungen ist es günstiger, diese Verteilung (äquivalent) in exponentieller Form zu schreiben, d.h.

$$p(y)=\exp\left[\sum_{i=2}^{n}g_i(y_{i-1},y_i)+\sum_{i=1}^{n}q_i(y_i)\right]-\ln(Z),\quad\text{mit}$$

$$Z=\sum_y\exp\left[\sum_{i=2}^{n}g_i(y_{i-1},y_i)+\sum_{i=1}^{n}q_i(y_i)\right].$$

Die unbekannten Parameter $\Theta=\big(q_i(k),\ 1\le i\le n,\ k\in K,\ \ g_i(k,k'),\ 2\le i\le n,\ k,k'\in K\big)$ in dieser Schreibweise entsprechen den logarithmierten ursprünglichen Parametern. Sie sollen anhand einer gegebenen Lernstichprobe der Zustandsfolgen $L=(y^1,y^2\dots y^l)$ nach dem Maximum-Likelihood Prinzip gelernt werden.

Geben Sie die Ableitungen der log-Likelihood Funktion nach den Parametern

$$\frac{\partial\ln p(L;\Theta)}{\partial q_i(k)}\quad\text{und}\quad\frac{\partial\ln p(L;\Theta)}{\partial g_i(k,k')}$$

an. Berücksichtigen Sie dabei, dass die Partition Funktion auch von Parametern abhängt. Schlussfolgern Sie daraus die notwendigen Bedingungen für die optimale Parameter.