MUSTERERKENNUNG,
6. SEMINAR – MAXIMUM-LIKELIHOOD PRINZIP

Aufgabe 1.
a) Die Wahrscheinlichkeitsverteilung einer skalaren Größe $x \in \mathbb{R}$ ist

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

mit reellen Parametern $\tau > 0$ und μ. Sie sollen nach dem Maximum-Likelihood Prinzip anhand einer Lernstichprobe $L = (x^1, x^2, \ldots, x^{|L|})$ angelernt werden. Wie ergeben sich daraus die gesuchten Größen?

b) Lösen Sie diese Aufgabe für die Wahrscheinlichkeitsverteilung

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

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 gesamten Lernstichprobe auch $P(L) = 0$ (und folglich nicht maximal). Offensichtlich müssen solche Werte von μ ausgeschlossen werden.

Aufgabe 2. 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 $L = \big((x^1, k^1), \ldots, (x^{|L|}, k^{|L|})\big)$. Seien die Streuungen σ_k bekannt. Man schätze μ mit Hilfe des Maximum-Likelihood Prinzips.

Aufgabe 3. Die Merkmale eines Objektes, welches sich in zwei Zuständen $k = 1,2$ befinden kann, sind Vektoren $x = (x_1, x_2) \in \mathbb{R}^2$. Die Wahrscheinlichkeitsverteilung ist

$$\begin{aligned} p(k{=}1) &= p(k{=}2), \\ p(x|k{=}1) &= C \cdot \exp\left[-\frac{(x_1-\mu_1)^2}{\sigma^2}\right], \\ p(x|k{=}2) &= C \cdot \exp\left[-\frac{(x_2-\mu_2)^2}{\sigma^2}\right], \end{aligned}$$

mit den Parametern $\mu_1, \mu_2, \sigma \in \mathbb{R}$.

a) Wie sieht die Klasse der Entscheidungsregeln für dieses Wahrscheinlichkeitsmodell aus (mit der Kostenfunktion $C(k,k') = \mathbb{I}(k{\neq}k')$)?

b) Seien die Parameter μ_1 und μ_2 unbekannt. Gegeben sei eine Lernstichprobe $L = \big((x^1,k^1),\dots,(x^{|L|},k^{|L|})\big)$. Finden Sie die unbekannten Parameter nach dem Maximum Likelihood Prinzip.

c) Finden Sie die Entscheidungsregel, die die Anzahl der Fehlklassifikationen auf der Lernstichprobe minimiert.

Aufgabe 4. Man betrachte „Mixtur der Gaussiane“ als Wahrscheinlichkeitsverteilung einer Zufallsvariable $x \in \mathbb{R}^n$:

$$p(x) = \sum_k w_k \cdot \mathcal{N}(x;\mu_k,\sigma) = \sum_k w_k \cdot \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{||x-\mu_k||^2}{2\sigma^2}\right],$$

mit $w_k \geq 0$ für alle k und $\sum_k w_k = 1$. Alle unbekannte Parameter der Mixtur (d.h. die Gewichte w_k, die Zentren μ_k und die Streuung σ) sollen anhand einer Lernstichprobe $L = (x^1, x^2 \dots x^{|L|})$ angelernt werden.

Hinweis: das Modell kann als eine Wahrscheinlichkeitsverteilung der Paare $p(k,x) = p(k) \cdot p(x|k)$ verstanden werden, in der die „Klassen“ k den Gaussianen entsprechen, d.h. $p(k) = w_k$ und $p(x|k) = \mathcal{N}(x;\mu_k,\sigma)$. Die Lernaufgabe besteht somit im unüberwachten Lernen der Parameter. Die Daten in der Lernstichprobe sind dabei unvollständig: nur x^l wird beobachtet, die Klasseninformationen k^l fehlen. Lösen Sie diese mittels Expectation-Maximization Algorithmus.