MUSTERERKENNUNG, 7. SEMINAR – GENERALIZED LINEAR MODELS

Aufgabe 1. Man betrachte die folgende Verbundwahrscheinlichkeitsverteilung für Paare $(x,y) \in \mathscr{X} \times \mathscr{Y}$. Gegeben sei eine Abbildung $\varphi : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}^d$, die jedem Paar (x,y) einen d-dimensionalen Vektor zuordnet. Die somit definierte Wahrscheinlichkeitsverteilung ist

$$p(x,y) = \frac{1}{Z} \exp\big(\langle \varphi(x,y), w \rangle\big)$$

($\langle \cdot \rangle$ bezeichnet das Skalarprodukt) mit einem Parametervektor $w \in \mathbb{R}^d$ und Normierungskonstante

$$Z = \sum_{xy} \exp\big(\langle \varphi(x,y), w \rangle\big).$$

a) Zeigen Sie, dass die Gaussche Wahrscheinlichkeitsverteilung (Einfachheit halber für eine Zufallsvariable $x \in \mathbb{R}$)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

ein Spezialfall davon ist.

Hinweis: Zeigen Sie, dass der Ausdruck unter dem Exponent als Skalarprodukt geschrieben werden kann. Wie ergeben sich somit die Abbildung φ und der Parametervektor w?

b) Man betrachte das überwachte Lernen des Parametervektors w. Gegeben sei eine Lernstichprobe von Paaren $L = \big((x^1,y^1),(x^2,y^2)\dots(x^l,y^l)\big)$. Leiten Sie den Gradient des log-Likelihoods $\partial \ln p(L)/\partial w$ ab. Wie ergeben sich daraus die Bedingungen für ein lokales Optimum des log-Likelihoods?

c) Man betrachte das unüberwachte Lernen, d.h. eine unvollständige Lernstichprobe $L = (x^1,x^2\dots x^l)$ steht zur Verfügung. Leiten Sie für diesen Fall den Gradient des log-Likelihoods ab.

d) Für das unüberwachte Lernen soll Expectation-Maximization Algorithmus verwendet werden. Welche „Abschätzungen“ müssen dabei im E-Schritt gemacht werden? Welche Optimierungsaufgabe entsteht im M-Schritt? Vergleichen Sie diese Vorgehensweise mit dem Gradientenanstieg-Verfahren aus **c)**.