MUSTERERKENNUNG, 8. SEMINAR – DISKRIMINATIVES LERNEN+

Aufgabe 1. Man betrachte das folgende Wahrscheinlichkeitsmodell mit zwei Klassen. Gegeben sei die a-priori Wahrscheinlichkeitsverteilung der Klassen $p(k)$, $k = 1,2$. Die bedingten Wahrscheinlichkeitsverteilungen der Beobachtungen $x \in \mathbb{R}$ sind

$$p(x|k) = \tau/2 \cdot \exp\big[-\tau|x-\mu_k|\big],$$

wobei der Parameter $\tau > 0$ für beide Klassen gleich ist. Leiten Sie die Formel für die a-posteriori Wahrscheinlichkeitsverteilung $p(k|x)$ ab.

Hinweis: Berücksichtigen Sie, dass die bedingten Wahrscheinlichkeitsverteilungen nicht überall differenzierbar sind. Somit muss man bei der Ableitung eine Fallunterscheidung machen.

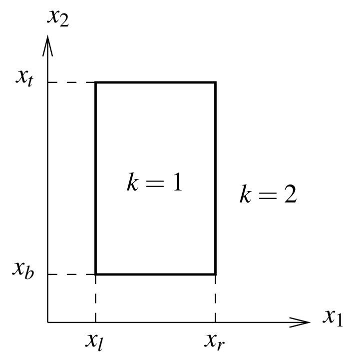
Aufgabe 2. (siehe das vorige Übungsblatt). Die Merkmale eines Objektes, welches sich in zwei Zuständen $k = 1,2$ befinden kann, sind Vektoren $x = (x_1, x_2) \in \mathbb{R}^2$. Die Wahrscheinlichkeitsverteilung ist

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

mit den Parametern $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$ und $\sigma \in \mathbb{R}$. Wie sieht die Klasse der Entscheidungsregeln für dieses Wahrscheinlichkeitsmodell aus? Bestimmen Sie die Vapnik-Chervonenkis Dimension dieser Klasse.

Aufgabe 3. Bestimmen Sie die Vapnik-Chervonenkis Dimension für die Entscheidungsregel, die im zweidimensionalen Raum ein achsenparalleles Rechteck bildet (siehe Skizze unten).

$$e(x) = \begin{cases} 1 & \text{wenn } x_l \leq x_1 \leq x_r \text{ und } x_b \leq x_2 \leq x_t \\ 2 & \text{sonst} \end{cases}$$



Aufgabe 4. (Generalized Linear Models). 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 mit der Normierungskonstante

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

a) Zeigen Sie, dass die Gaussche Wahrscheinlichkeitsverteilung (Einfachheit halber für eine eindimensionale 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 $\langle \varphi(x), w \rangle$ geschrieben werden kann. Wie ergeben sich daraus die Abbildung φ und der Parametervektor w?

b) Man betrachte das überwachte Lernen des Parametervektors w nach dem Maximum-Likelihood Prinzip. Gegeben sei eine Lernstichprobe $L = \big((x^1,y^1),(x^2,y^2)\ldots(x^l,y^l)\big)$. 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) Wir modifizieren das Modell, indem wir die ursprüngliche Formel als bedingte Wahrscheinlichkeitsverteilung verstehen, d.h. das Modell ist jetzt $p(x,y) = p(x) \cdot p(y|x)$ mit

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

und einer beliebigen Wahrscheinlichkeitsverteilung $p(x)$. Die Normierungskonstante ist jetzt x-spezifisch, d.h.

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

und sorgt dafür, dass die bedingten Wahrscheinlichkeitsverteilungen für alle x auf 1 normiert sind. Leiten Sie für diesen Fall den Gradient des log-Conditional Likelihoods ab und vergleichen Sie ihn mit **b)**.