Übungen zur Vorlesung

Mustererkennung und Neuronale Netze

Wintersemester 2000/2001

Institut für Informatik III
Prof. J. Buhmann
Dipl.-Inform. M. Braun
Tel.: (02 28) 73–42 19
Fax: (02 28) 73–43 82
Email: {jb,braunm}@cs.uni-bonn.de
WWW: http://www-dbv.cs.uni-bonn.de

Blatt 9: Schranken für große Abweichungen und Klassifikationsfehler bei linearer Klassifikation

Abgabe Montag, 29. Januar 2001, in der Vorlesung

11. *Schranken für große Abweichungen:* In der Vorlesung wurden einige Abschätzungen für die Wahrscheinlichkeit einer großen Abweichung $P\{X - EX \geq \varepsilon\}$ einer Zufallsvariable X vorgestellt.

    In dieser Übung sollen die Abschätzungen qualitativ verglichen werden. Im einzelnen betrachtet werden:

    (1) Bernoulliverteilung $B(p)$ zu $0 \leq p \leq 1$. X ist $B(p)$ verteilt, wenn $X \in \{0, 1\}$ und $P\{X = 1\} = p$, $P\{X = 0\} = 1 - p$.

    (2) Normalverteilt $N(\mu, \sigma)$, mit Parametern $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, Dichte

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

    Bearbeiten Sie nun folgende Aufgaben.

    (a) Bestimmen Sie $P\{X - EX \geq \varepsilon\}$ für $B(p)$ verteiltes X. Die entsprechende Formel für die Normalverteilung läßt sich nicht geschlossen darstellen. Sie können jedoch Näherungen durch die Matlab Funktion `erf` berechnen.
    Plotten Sie die tatsächlichen Wahrscheinlichkeiten von $B(p)$ und $N(0, \sigma^2)$ gegen Abschätzungen der Markov- und Chebyshev-Ungleichung. Variieren Sie die Parameter. **(3P)**

    (b) Wir gehen nun zu Summen von unabhängig, identisch verteilten Zufallsvariablen über. Sei $X_i \sim B(p)$ ($i \in \mathbb{N}$).
    Die Verteilung von $S_n := \sum_{i=1}^n X_i$ ist gegeben durch die Binomialverteilung: $P\{S_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}$. Bestimmen Sie $E(S_n)$.
    Plotten Sie die tatsächlichen Wahrscheinlichkeiten $P\{S_n - ES_n \geq \varepsilon\}$, sowie die Abschätzungen durch die Hoeffdings- und die Bernstein-Ungleichung. **(3P)**

12. *Klassifikationsfehler von linearer Klassifikation:* In der Vorlesung wurde folgendes Resultat zitiert: Sei $\mathcal{H}$ die Menge aller Hyperebenen. Falls $n \geq d$ und $2d/n \leq \varepsilon \leq 1$, gilt für die Hyperebene $\widehat{\phi}$, welche den empirischen Fehler R_{emp} minimiert,

    $$P\{R(\widehat{\phi}) > \inf_{\phi \in \mathcal{H}} R(\phi) + \varepsilon\} \leq e^{2d\varepsilon} \left(2\binom{n}{d} + 1\right) e^{-n\varepsilon^2/2}, \tag{1}$$

    Außerdem gilt, falls $n \geq d$,

    $$E\{R(\widehat{\phi}) - \inf_{\phi \in \mathcal{H}} R(\phi)\} \leq \sqrt{\frac{2}{n}((d+1)\log n + (2d+2))}. \tag{2}$$

    Diese Abschätzungen gelten nicht für die Hyperebene, welche den mittleren quadratischen Fehler minimiert (siehe Aufgabe 10). Das Verhalten dieser Hyperebene soll jetzt mit obigen Gleichungen verglichen werden.

(a) Generieren Sie einen Datensatz wie in Aufgabe 3 (a) mit $P(\omega) = \frac{1}{2}$, $\omega \in \{1, 2\}$, $\mu_1 = (-5, 0)^t$, $\mu_2 = (3, 3)$, $\Sigma_1 = \Sigma_2 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$. Die optimale Diskriminanzfläche ϕ^* ist dann eine Linie (wieso?). Bestimmen Sie ϕ^*. Bestimmen Sie durch Sampling den erwarteten Klassifikationsfehler $\inf_{\phi \in \mathcal{H}} R(\phi) =: R^*$. **(3P)**

(b) Nun soll das Verhalten der Widrow-Hoff-Lernregel (siehe Aufgabe 10) mit den Gleichungen (1) und (2) verglichen werden. Ziehen Sie hierzu einen Datensatz, lernen Sie mittels Widrow-Hoff auf einem Teil der Daten und schätzen Sie den Klassifikationsfehler $R(\widehat{\phi})$ auf den restlichen Daten ab. Tragen Sie die $R(\widehat{\phi}) - R^*$ in ein Histogramm ein. Wiederholen Sie den Vorgang für $k = 100$ Durchläufe.

Bestimmen Sie nun angenäherte Werte $\tilde{P}\{R(\widehat{\phi}) - R^* \geq \varepsilon\}$ und $\tilde{E}\{R(\widehat{\phi}) - R^*\}$ aus dem Histogramm und vergleichen Sie diese mit Gleichungen (1) und (2). **(5P)**