Übungen zur Vorlesung

Mustererkennung und Neuronale Netze

Wintersemester 2000/2001

Institut für Informatik III
Prof. J. Buhmann
Dipl. Inform. T. Zöller
Dipl. Inform. L. Hermes
Tel.: (02 28) 73–41 02
Fax: (02 28) 73–43 82
Email: {jb,zoeller,hermes}@cs.uni-bonn.de
WWW: http://www-dbv.cs.uni-bonn.de

Blatt 3: Schätzen von multivariaten Normalverteilungen, Bayes–Klassifikator (Teil 2)

Abgabe Montag, den 20. November, in der Vorlesung

4. *Schätzen von multivariaten Normalverteilungen:* Es soll eine Funktion zum Schätzen einer Normalverteilung aus den empirschen Daten implementiert werden. Schreiben Sie dazu eine Funktion

```
function [mu,Sigma] = GAUSS(X),
```

indem Sie für empirische Daten $\{x_i : 1 \leq i \leq n\}$ die erwartungstreuen Schätzer $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ und $\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^t$ verwenden.

(a) Vergleichen Sie die Ergebnisse mit den Matlab–Funktionen `mean()` und `cov()` auf Basis der in Aufg. 2 b) erstellten Daten. **(1P)**

(b) Plotten Sie den empirischen Schätzfehler $\|\hat{\mu} - \mu\|$ in Abhängigkeit von der Samplegröße n für die Parameter aus Aufg. 2 b). Plotten Sie zum Vergleich $\det(\Sigma)/\sqrt{n}$ (Warum?). **(2P)**

(c) Überprüfen Sie die Erwartungstreue: generieren Sie 20 Datenpunkte gemäß $\mathcal{N}(0, 1)$ und schätzen Sie Mittelwert und Varianz. Wiederholen Sie diesen Vorgang 10000 mal und stellen Sie die Verteilung der geschätzten Mitelwerte und Varianzen mittels der Matlab Funktion `hist()` dar. Schätzen Sie den 'mittleren' empirischen Mittelwert und die 'mittlere' empirische Varianz mittels `mean()`. **(2P)**

(d) *Freiwillige Aufgabe:* Untersuchen Sie empirisch den Maximum–Likelihood–Schätzer für die (Ko)varianz $\breve{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^t$ auf seine Erwartungstreue. **(1P)**

5. *Bayes–Klassifikator:*

In Ergänzung zur *Bayes–Klassifikator*–Aufgabe vom letzten Übungsblatt (Aufgabe 3) bearbeiten Sie folgende weitere Teilaufgabe:

(a) Verwenden Sie $n \in \{10, 20, 50, 100, 200, 500, 1000\}$ zufällig ausgewählte Trainingsdaten, um mittels der Funktion `GAUSS()` die Parameter μ_ω und Σ_ω zu schätzen. Schätzen Sie außerdem die a priori Klassenwahrscheinlichkeiten mittels $\hat{p}_\omega = \frac{1}{n}|x : \omega^{\text{true}}(x) = \omega|$. Schätzen Sie den Klassifikationsfehler mit den nicht verwendeten 9000 Datenpunkten. Plotten Sie den Klassifikationsfehler in Abhängigkeit von n auf einer logarithmischen Skala. **(2P)**