Maschinelles Lernen
Übungsblatt 5

Dr. Steffen Rendle, Christoph Freudenthaler
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

23. November 2010
Abgabe bis 3. Dezember

Diskriminanzanalyse I

Wissenschaftler haben die Böden von Iowa, welche ein bestimmtes Bakterium enthalten (Klasse 1), mit andere Böden, die es nicht enthalten (Klasse 2), verglichen. Dabei haben sie die Variablen x_1 (pH-Wert) und x_2 (Stickstoffgehalt) beobachtet. Die Anzahl der Instanzen pro Klasse, der Mittelwert der Vektoren und die Kovarianzmatrizen für die zwei Bodenarten seien wie folgt gegeben:

$$n_1 = 13, \qquad n_2 = 10$$
$$\mathbf{m}_1 = \begin{pmatrix} 7.8 \\ 45 \end{pmatrix}, \qquad \mathbf{m}_2 = \begin{pmatrix} 5.9 \\ 20.8 \end{pmatrix}$$
$$\mathbf{S}_{W1} = \begin{pmatrix} 0.5 & 4.5 \\ 4.5 & 147.2 \end{pmatrix}, \quad \mathbf{S}_{W2} = \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 24.2 \end{pmatrix}$$

a) Stellen Sie die Diskriminanzfunktionen für die beiden Klassen auf.

b) Ordnen Sie die Beobachtung $x = \begin{pmatrix} 6 & 52.5 \end{pmatrix}^T$ einer der beiden Klassen zu.

c) Handelt es sich hier um lineare oder quadratische Diskriminanzanalyse? Nennen Sie die Unterschiede zwischen LDA und QDA.

Diskriminanzanalyse II

Bei der Diskriminanzanalyse wird angenommen, dass, gegeben der Kategorie $Y \in \{0, 1\}$, die Verteilung der erklärenden Variablen X eine Normalverteilung ist.

a) Wiederholen Sie die Annahmen der Diskriminanzanalyse und geben Sie die bedingten Wahrscheinlichkeiten $p(X|Y)$ für beide Klassen $Y \in \{0, 1\}$ an. Erklären Sie, warum die Diskriminanzanalyse zur Klasse der prototypischen Methoden gezählt wird.

b) Die Maximum-Likelihood-Schätzer für die Wahrscheinlichkeit $p(Y = k) = \pi_k$ der beiden Klassen $k \in \{0, 1\}$ bzw. deren bedingte Erwartungswerte μ_k sind

$$\hat{\pi}_k = \frac{n_k}{n}, \qquad \hat{\mu}_k = \frac{1}{n_k} \sum_{\{i:y_i=k\}} x_i$$

Leiten Sie die beiden Schätzfunktionen $\hat{\pi}_k$ und $\hat{\mu}_k$ aus den Modellannahmen her.

Hinweis: Verwenden Sie als Ausgangspunkt zur Bestimmung der Schätzfunktionen die Likelihoodfunktion der Daten und maximieren sie diese *jeweils* nach den gesuchten Parameter μ_k und π_k. Die Likelihoodfunktion der Daten D vom Unfang $|D| = n$ lautet

$$p(D|\mu_1, \mu_2, \pi_1, \pi_2, \Sigma_1, \Sigma_2) = \prod_{i=1}^{n} p(x_i|y_i, \mu_1, \mu_2, \pi_1, \pi_2, \Sigma_1, \Sigma_2)p(y_i)$$

LDA vs. QDA vs. Logistische Regression

a) Offensichtlich erzeugt die quadratische Diskriminanzanalyse nicht-lineare Entscheidungsgrenzen. Warum ist es manchmal notwendig, anstelle einfacher linearer Entscheidungsgrenzen, komplexere Entscheidungsgrenzen zu erlauben? Welche Nachteile hat die Verwendung komplexerer Entscheidungsgrenzen?

b) Aus welchem Grund erzeugt die QDA quadratische Entscheidungsgrenzen, die LDA aber nicht? Wie könnte man die Daten transformieren, sodass anstelle einer komplexeren QDA eine einfachere LDA zur Unterscheidung zwischen den beiden Klassen $Y \in \{0, 1\}$ ausreicht? Können so die Nachteile der komplexerer QDA beseitigt werden?

c) Betrachten sie die Modellannahmen der logistischen Regression und der Diskriminanzanalyse. Benennen Sie die Modellannahmen und beschreiben Sie deren Unterschiede. Welche der beiden Methoden, d.h. welche Modellannahmen, halten sie für die allgemeineren? Begründen Sie. Ist es besser möglichst konkrete Modellannahmen zu treffen oder doch eher möglichst allgemeine?