Dr. Katharina Best
Sommersemester 2011
19. Mai 2011

Reader – Teil 6: Diskriminanzanalyse

WiMa-Praktikum

Im gegensatz zur Clusteranalyse haben wir bei der Diskriminanzanalyse die Klassen vorgegeben. Was uns fehlt, ist jedoch die Klasseneinteilung auf Grund des Merkmalvektors. Solche Probleme werden in der Medizin zur Diagnostik gestellt, aber auch zur Rasterfahnung von Personen nach Terrorist/kein Terrorist bzw. allgemein im Data Mining.

Insgesamt gehören sie zum Themengebiet der Mustererkennung (engl. pattern recognition). Da die Merkmalsausprägungen, die die Grundlage der Klassifikation klassifiziert bilden, zufälligen Schwankungen unterliegen, haben wir es mit einem statistischen Problem zu tun. Viele Klassifikationsverfahren stammen aus der Informatik.

Prinzipiell wollen wir uns Mannigfaltigkeiten, *Klassifikatoren* genannt, im Merkmalsraum definieren, die möglichst gut trennen und trotzdem möglichst glatt sind. Oft werden das Hyperebenen sein. Mit ihrer Hilfe zerlegen wir den Stichprobenraum in Klassengebiete. Dazu benötigen wir aber zuerst eine Stichprobe, bei der die Klassenzugehörigkeit bekannt ist und die zur Bestimmung des Klassifikators herangezogen werden. Diese Stichprobe nennen wir *Lernstichprobe*. Diese muss von der danach zu klassifizierenden Stichprobe getrennt gehalten werden.

1 Grundidee

Haben wir m Klassen $K_1, \ldots, K_m$, so benötigen wir m Klassengebiete $G_1, \ldots, G_m$ des Mermalraums S. Wir ordnen ein Objekt ω, für welches wir den Merkmalsvektor x beobachten mittels der Entscheidungsregel

$$\omega \in K_i \qquad \Longleftrightarrow \qquad x \in G_i \tag{1}$$

zu. Dazu definieren wir uns eine Zufallsvariable A mit $\mathbb{P}(A = i) = \mathbb{1}_{K_i}$, d. h. wir bekommen genau den Index der Gruppe zu der ω gehört.

Hier stellen wir nur die zwei wichtigsten vor.

2 Bayessches Vorgehen

Wir gehen von einer a priori Verteilung der Klassen aus, also einer Zähldichte p_i mit $1 \leq i \leq m$. Weiter benötigen wir die bedingte Verteilung des Merkmalvektors

$$\mathbb{P}(X \in G_i | A = j). \tag{2}$$

Wir wollen dann Aussagen über die a posteriori Wahrscheinlichkeiten

$$\mathbb{P}(A = j | X) \tag{3}$$

treffen. Dazu können wir den Satz von Bayes anwenden. Ein Objekt wird dann der Klasse zugeordnet, für die es die grösste a postesteriori Wahrscheinlichkeit hat, also ist

$$\hat{k} = \text{argmax}_{1 \leq j \leq m} \mathbb{P}(A = j | X) \tag{4}$$

der Schätzer für den Gruppenindex.

3 Lineare Diskriminanzfunktion

Wir gehen davon aus, dass die Merkmalsvektoren Normalverteilungen folgen, welche identische Kovarianzmatrizen haben und unterschiedliche Erwartungswerte, d. h. dass wir eine Dichte für die Verteilung von X gegeben der Klassenzugehörigkeit haben. Eine Beobachtung wird dann der Klasse zugeordnet, bezüglich derem Erwartungswert die Mahalanobis-Distanz des Merkmalsvektor am kleinsten ist, also ist

$$\hat{k} = \text{argmax}_{1 \leq j \leq m} \delta_S(X, \mu_j) \tag{5}$$

der entsprechende Schätzer.

4 Diskussion

Fragen die wir uns immer stellen müssen, sind die nach der richtigen Klassifizierung.

- Welcher Anteil der Merkmalsvektoren kann allein vom Verfahren her nicht richtig klassifiziert werden?
- Sind die Fehlklassifizierungen symmetrisch oder gibt es eine besonders schlecht abschneidende Klasse?
- Welche Folgen hat eine Fehlklassifizierung? Wollen wir eine besondere davon explizit minimieren?