

Statistik IV für Nebenfachstudierende

5 Diskriminanzanalyse I

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Motivation

- Die Grundgesamtheit zerfällt in k verschiedene Gruppen oder Klassen.

$$Y \in \{1, \dots, k\}$$

- Die Gruppen / Klassen unterscheiden sich bezüglichlicher Merkmale.
- Ziel ist es von den Merkmalen auf die Klasse zu schließen.
 - Für jedes Objekt verfügbar: $\underline{x}^\top = (x_1, \dots, x_p)$
 - Gesucht: Klasse des Objekts

Beispiele

Synonyme

- Diskriminanzanalyse
- Klassifikation
- Pattern recognition
- Supervised Learning

Beispiel: Diagnostischer Test

- Sensitivität:
- Spezifität:

Wichtige Größen

- *a priori*-Wahrscheinlichkeiten

$$p(r) = P(Y = r), r = 1, \dots, k,$$

- *a posteriori*-Wahrscheinlichkeiten

$$P(r | \underline{x}) = P(Y = r | \underline{x}), r = 1, \dots, k,$$

- *Verteilung der Merkmale*, gegeben die Klasse, bestimmt durch die Dichten

$$f(\underline{x} | 1), \dots, f(\underline{x} | k).$$

- *Mischverteilung* der Population (bzw. die unbedingte Verteilung von \underline{x}):

$$f(\underline{x}) = p(1)f(\underline{x} | 1) + \dots + p(k)f(\underline{x} | k).$$

Bayes Zuordnung

- Gesucht: Zuordnungsregel

$$\delta : \mathbb{R}^p \longrightarrow \{1, \dots, k\}$$

$$\underline{x} \longrightarrow \delta(\underline{x})$$

- Bayes Zuordnung:

$$\delta^*(\underline{x}) = r \iff P(Y = r | \underline{x}) = P(r | \underline{x}) = \max_{i=1, \dots, k} P(i | \underline{x})$$

Satz von Bayes

- Für die Bayes Zuordnung brauchen wir $P(Y = r | \underline{x})$
- Wenn $f(x | Y = r)$ bekannt, nutzt man dazu den Satz von Bayes:

Satz von Bayes

$$P(r | \underline{x}) = \frac{f(\underline{x} | r)p(r)}{f(\underline{x})} = \frac{f(\underline{x} | r)p(r)}{\sum_{i=1}^k f(\underline{x} | i)p(i)}$$

Fehlklassifikationswahrscheinlichkeiten

- Betrachtet wird eine *feste* Zuordnung

$$\delta : \mathbb{R}^p \longrightarrow \{1, \dots, k\}$$

- Zufällige Beobachtung (Y, \underline{x})

Globale Fehlerrate

- Globale Fehlerrate (Gesamt-Fehlerrate)
entspricht Fehlklassifikations-Wahrscheinlichkeit

$$\varepsilon = P(\delta(\underline{x}) \neq Y)$$

Fehlerrate bei gegebenen \underline{x}

- Wahrscheinlichkeit einer Fehlklassifikation, gegeben der feste Merkmalsvektor \underline{x}

$$\begin{aligned}\varepsilon(\underline{x}) &= P(\delta(\underline{x}) \neq Y \mid \underline{x}) \\ &= 1 - P(\delta(\underline{x}) = Y \mid \underline{x}) \\ &= 1 - P(\delta(\underline{x}) \mid \underline{x}).\end{aligned}$$

Individuelle Fehlerrate

- Verwechslungswahrscheinlichkeit oder individuelle Fehlerrate

$$\begin{aligned}\varepsilon_{rs} &= P(\delta(\underline{x}) = s \mid Y = r) \\ &= \int_{\underline{x}: \delta(\underline{x})=s} f(\underline{x} \mid r) d\underline{x}, \quad r \neq s\end{aligned}$$

Fehlerrate für Klasse r

- Wahrscheinlichkeit einer Fehlklassifikation, gegeben das Objekt entstammt der Klasse r

$$\begin{aligned}\varepsilon_r &= P(\delta(\underline{x}) \neq r \mid Y = r) \\ &= \sum_{s \neq r} \varepsilon_{rs}\end{aligned}$$

Es gilt der Zusammenhang

Für die globale Fehlerrate gilt:

$$\begin{aligned}\varepsilon &= P(\delta(\underline{x}) \neq Y) = \sum_{r=1}^k P(\delta(\underline{x}) \neq r \mid Y = r)p(r) = \sum_{r=1}^k \varepsilon_r p(r) \\ &= \sum_{r=1}^k \sum_{s \neq r} \varepsilon_{rs} p(r) \\ \varepsilon &= P(\delta(\underline{x}) \neq Y) = \int P(\delta(\underline{x}) \neq Y \mid \underline{x}) f(\underline{x}) d\underline{x} = \int \varepsilon(\underline{x}) f(\underline{x}) d\underline{x}\end{aligned}$$

Verbindung zur Bayes-Zuordnung

Die Bayes-Zuordnung

$$\delta^*(\underline{x}) = r \iff P(r | \underline{x}) = \max_{i=1, \dots, k} P(i | \underline{x})$$

minimiert die Gesamtfehlerrate ε .

Beispiel

