3.Tutorium Multivariate Verfahren

- Diskriminanzanalyse -

Hannah Busen:

11.05.2015 und 18.05.2015

Nicole Schüller:

12.05.2015 und 19.05.2015

Institut für Statistik, LMU München

Gliederung

Gliederung

Diskriminanzanalyse als Klassifikationsverfahren

- Die betrachtete Grundgesamtheit zerfällt in k disjunkte Populationen mit Indikator $y \in \{1, \ldots, k\}$
- Jedes Individuum soll in eine der k Klassen zugeordnet werden
- Für jedes Individuum werden zusätzliche Merkmale erhoben
  Man beobachtet den Merkmalsvektor $\mathbf{x}^\top = (x_1, \ldots, x_p)$
- Ziel: von **x** auf y schließen
- Beispiel: **Kreditscoring**
  - Kreditnehmer sollen in die beiden Klassen „kreditwürdig“ und „nicht kreditwürdig“ eingeteilt werden
  - beobachtete Merkmale: Alter, Beschäftigungsart/-dauer, Einkommen

Diskriminanzanalyse als Klassifikationsverfahren

- I.d.R liegen für n Individuen die wahre Klassenzugehörigkeit $y_i \in \{1, \ldots, k\}$, sowie die Merkmalsvektoren $\mathbf{x}_i$, $i = 1, \ldots, n$ vor
- Anhand dieser Daten $(y_i, \mathbf{x}_i^\top)$, $i = 1, \ldots, n$ wird eine Entscheidungsfunktion $\delta(\mathbf{x})$ geschätzt, mit der die Individuen mittels ihrer Merkmalsvektoren einer Klasse zugeordnet werden
- Diese Entscheidungsfunktion $\delta(\mathbf{x})$ kann auf weitere Individuen angewendet werden, deren wahre Klassenzugehörigkeit unbekannt ist
  $\Rightarrow$ **Fehlklassifikation möglich**

Gliederung

Fehlerraten

Gegeben: feste Zuordnungsregel $\delta(\mathbf{x})$, Zufallsvektor $(y, \mathbf{x}^\top)$

- Globale Fehlklassifikationswahrscheinlichkeit:

  $\varepsilon = P(\delta(\mathbf{x}) \neq y)$

- Fehlklassifikation, gegeben $\mathbf{x}$:

  $\varepsilon(\mathbf{x}) = P(\delta(\mathbf{x}) \neq y|\mathbf{x}) = 1 - P(\delta(\mathbf{x}) = y|\mathbf{x})$

- Verwechslungswahrscheinlichkeit:

  $\varepsilon_{rs} = P(\delta(\mathbf{x}) = s|y = r)$

- Fehlklassifikation, gegeben das Objekt ist aus Klasse r:

  $\varepsilon_r = P(\delta(\mathbf{x}) \neq r|y = r) = \sum_{s \neq r} \epsilon_{rs}$

Gliederung

Bayes-Zuordnung

- Zuordnung zur Klasse mit größter a-posteriori-Wahrscheinlichkeit
- Vergleiche: $P(y=1|\mathbf{x})\ldots P(y=k|\mathbf{x})$
- Bayes-Zuordnungsregel:

$$\delta^*(\mathbf{x}) = r \Leftrightarrow P(y=r|\mathbf{x}) = \max_{j=1,\ldots,k} P(y=j|\mathbf{x})$$

- die a-posteriori-Wahrscheinlichkeiten $P(y=r|\mathbf{x})$ erhält man über den Satz von Bayes
- die a-priori-Wahrscheinlichkeiten $P(y=r)=p(r)$ werden aus den Daten geschätzt
- **Merke:** Bayes-Zuordnung minimiert die Gesamt-Fehlerrate ε

Maximum-Likelihood-Zuordnung

- Zuordnung zu der Klasse mit maximaler Likelihood
- Vergleiche: $f(\mathbf{x}|y=1)\ldots f(\mathbf{x}|y=k)$
- ML-Zuordnung:

$$\delta_{ML}(\mathbf{x}) = r \Leftrightarrow f(\mathbf{x}|y=r) = \max_{j=1,\ldots,k} f(\mathbf{x}|y=j)$$

- ML-Zuordnung entspricht Bayes-Zuordnung bei gleichen a-priori-Wahrscheinlichkeiten!
  $\Rightarrow$ ML-Zuordnung ist Spezialfall der Bayes-Zuordnung!

Kostenoptimale Zuordnung

- Definiere die Kostenfunktion $c_{ij} = c(i, j)$
- $c_{ij} \hat{=}$ Kosten einer Zuordnung eines Individuums aus Klasse i in Klasse j Es gilt: $c_{ij} \geq 0$, $c_{ii} = 0, \; \forall i, j$
- Zuordnung:

$$\delta(\mathbf{x}) = r \Leftrightarrow \sum_{i=1}^{k} P(y = i|\mathbf{x})c_{ir} = \min_{j=1,\dots,k} \sum_{i=1}^{k} P(y = i|\mathbf{x})c_{ij}$$

- $r(\mathbf{x}) = P(y = i|\mathbf{x})c_{ij} \hat{=}$ Risiko bei Zuordnung in Klasse j
- Spezialfälle:
  (a) $c_{ij} = c \hat{=}$ Bayes-Zuordnung
  (b) $c_{ij} = \frac{c}{p(i)} \hat{=}$ ML-Zuordnung

Gliederung

Bayes-Zuordnung mit Diskriminanzfunktion

- Definiert wird die Diskriminanzfunktion $d_r(\mathbf{x}) = P(y = r|\mathbf{x})$ für die r-te Klasse
- Das Individuum mit Merkmalsvektor $\mathbf{x}$ wird somit der Klasse zugeordnet, für die die zugehörige Diskriminanzfunktion maximal ist
- **Äquivalente** Diskriminanzfunktionen, die auch Bayes-Zuordnung bewirken:
  (a) $d_r(\mathbf{x}) = P(y = r|\mathbf{x})$
  (b) $d_r(\mathbf{x}) = f(\mathbf{x}|r) \cdot p(r)$
  (c) $d_r(\mathbf{x}) = \log(f(\mathbf{x}|r)) + \log(p(r))$

Gliederung

Diskriminanzfunktion unter NV-Annahme

- Für jede Dichte $f(\mathbf{x}|y = r)$ wird eine multivariate Normalverteilung mit Erwartungswert $\boldsymbol{\mu}_r$ und Kovarianzmatrix $\boldsymbol{\Sigma}_\mathbf{r}$ angenommen:

$$x|\text{r-te Klasse} \sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_\mathbf{r})$$

- Dichte in Klasse r:

$$f(\mathbf{x}|r) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_\mathbf{r}|^{1/2}} \cdot \exp(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r})^\top {\boldsymbol{\Sigma}_\mathbf{r}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r}))$$

- Diskriminanzfunktion (vgl. Folie 13):

$$\begin{aligned} d_r(\mathbf{x}) &= \log(f(\mathbf{x}|r)) + \log(p(r)) \\ &\propto -\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r})^\top {\boldsymbol{\Sigma}_\mathbf{r}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r}) - \tfrac{1}{2}\log(|\boldsymbol{\Sigma}_\mathbf{r}|) + \log(p(r)) \end{aligned}$$

Fallunterscheidungen

1. $\mathbf{x}|$r-te Klasse $\sim N(\boldsymbol{\mu}_r, \sigma^2 \mathbf{I})$ **unkorreliert!**

$$\begin{aligned} d_r(\mathbf{x}) &= -\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu_r})^\top (\mathbf{x} - \boldsymbol{\mu_r}) + \log(p(r)) \\ &= \mathbf{a_r}^\top \mathbf{x} + a_{r0} \end{aligned}$$

2. $\mathbf{x}|$r-te Klasse $\sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma})$ **unabhängig von r!**

$$\begin{aligned} d_r(\mathbf{x}) &= -\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu_r})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu_r}) + \log(p(r)) \\ &= \mathbf{a_r}^\top \mathbf{x} + a_{r0} \end{aligned}$$

Lineare Funktion trennt die Klassen!

$\Rightarrow$ **Lineare** Diskriminanzanalyse!

Fallunterscheidungen

3. $\mathbf{x}|$r-te Klasse $\sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_\mathbf{r})$ $\rightarrow$ Verschiedene Kovarianzen in jeder Klasse!

$$\begin{aligned} d_r(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r})^\top \boldsymbol{\Sigma}_\mathbf{r}{}^{-1}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{r}) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_\mathbf{r}|) + \log(p(r)) \\ &= \mathbf{x}^\top \mathbf{A}_r \mathbf{x} + \mathbf{a}_r^\top \mathbf{x} + a_{r0} \end{aligned}$$

$\rightarrow$ Keine Terme vernachlässigbar!

Quadratische Funktion trennt die Klassen!

$\Rightarrow$ **Quadratische** Diskriminanzanalyse!