Statistik IV für Nebenfachstudierende

5 Diskriminanzanalyse II

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Bayes Zuordnung mit Diskriminanzfunktion

Wiederholung Bayes Zuordnung:

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

- Bayes Zuordnung:

$$\delta^*(\underline{x}) = r \Longleftrightarrow P(Y = r \mid \underline{x}) = P(r \mid x) = \max_{i=1,\ldots,k} P(i \mid \underline{x})$$

Diskriminanzfunktionen

Betrachte “Scores“:

$$\delta^*(\underline{x}) = r \iff d_r(\underline{x}) = \max_{i=1,\ldots,k} d_i(\underline{x})$$

- Wenn $d_r(\underline{x}) = P(r \mid x)$ resultiert Bayes Zuordnung.
- Bei zwei Klassen:

Bayes Zuordnung resultiert aus...

a) $d_r(\underline{x}) = P(r \mid \underline{x})$,

b) $d_r(\underline{x}) = f(\underline{x} \mid r)p(r)/f(\underline{x})$.

c) $d_r(\underline{x}) = f(\underline{x} \mid r)p(r)$.

d) $d_r(\underline{x}) = \log(f(\underline{x} \mid r)) + \log(p(r))$.

Bild: Bayes Zuordnung

Bild: Keine Bayes Zuordnung

Bild: Bayes bei ungleichen prioris

Bild: Bayes auch hier...

Maximum Likelihood Regel

Beispiel:

- Bayes Zuordnung diskriminiert nicht mehr.

Maximum Likelihood Regel

ML Zuordnung:

$$\delta_{\mathsf{ML}}(\underline{x}) = r \Longleftrightarrow f(\underline{x} \mid r) = \max_{i=1,\ldots,k} f(\underline{x} \mid i)$$

- Ist nicht optimal bzgl. Fehlerrate
- Folgt ML Prinzip, die Klassen sind wie Parameter.
- Wähle die Klasse, welche am ehesten für das Zustandekommen von $\underline{x}$ verantwortlich ist.

Entscheidungstheoretischer Zugang

δ entscheidet auf Grund von $\underline{x}$.

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

- Kosten für Entscheidungen (Fehlzuordnungen)

  $c(i,j) = c_{i,j}$

- Annahmen:
  - $c_{ii} = 0 \quad \forall i$
  - $c_{ij} \geq 0 \quad \forall i,j$

Für zufällige Beobachtungen $(Y, \underline{x})$ und feste Regel $\delta()$ wird

$$c(\ Y,\ \delta(\underline{x})\ )$$

eine Zufallsvariable.

- Entscheidungstheorie zur Regel $\delta()$ betrachtet zu erwartende Kosten (Schaden) als Risiko.
- Das Gesamtrisiko ist

$$R_\delta = E_{Y,\underline{x}}\left(c(Y, \delta(\underline{x}))\right)$$

Risiko-Maße

- Bedingtes Risiko, gegeben $\underline{x}$

$$r(\underline{x}) = \sum_{i=1}^{k} c_{i,\delta(\underline{x})} P(i \mid \underline{x}).$$

- Individuelles Risiko

$$r_{ij} = c_{ij} P(\delta(\underline{x}) = j \mid Y = i) = c_{ij} \int_{\underline{x}:\delta(\underline{x})=j} f(\underline{x} \mid i) d\underline{x}$$

- Risiko gegeben Klasse i

$$r_i = \sum_{j=1}^{k} r_{ij}.$$

Gesamt-Risiko

Das gesamte Bayes Risiko, d.h. die zu erwartenden Kosten, lassen sich darstellen durch

$$R_\delta = E(c(Y, \delta(\underline{x}))) = \sum_{i=1}^{k} r_i p(i) = \int r(\underline{x}) f(\underline{x}) d\underline{x}.$$

Bayes Zuordnung mit Kostenfunktion

Bei Vorliegen des Beobachtungsvektors $\underline{x}$ wird das

Bayes–Risiko minimiert durch die Zuordnung

$$\delta^*(\underline{x}) = r \Longleftrightarrow \sum_{i=1}^{k} P(i \mid \underline{x}) c_{ir} = \min_{j=1,\ldots,k} \sum_{i=1}^{k} P(i \mid \underline{x}) c_{ij}$$

Mit der Diskrimininanzfunktion

$$d_r(\underline{x}) = -\sum_{i=1}^{k} P(i \mid \underline{x}) c_{ir}$$

erhält man

$$\delta^*(\underline{x}) = r \Longleftrightarrow d_r(\underline{x}) = \max_{i=1,\ldots,k} d_i(\underline{x})$$

Spezialfall: fixe Kosten

Fixe Kosten:

- $c_{ij} = 0$ für $i = j$
- $c_{ij} = c$ für $i \neq j$

- Bayes Zuordnung:

Umgekehrt proportionale Kostenfunktion

Kosten invers zu $p(i)$:

- $c_{ij} = 0$ für $i = j$
- $c_{ij} = \frac{c}{p(i)}$ für $i \neq j$

- Bayes Zuordnung:

Zuordnung unter Normalverteilung

$$\underline{\underline{X}}|Y = r \quad \sim \quad N(\underline{\mu}_r\,,\ \boldsymbol{\Sigma}_r) \quad \text{für} \quad r = 1, \ldots, k$$

Zuordnung unter Normalverteilung

Normalverteilte Merkmale in den Klassen:

$$\underline{\underline{X}}|Y = r \quad \sim \quad N(\underline{\mu}_r\,,\ \boldsymbol{\Sigma}_r) \quad \text{für} \quad r = 1, \ldots, k$$

d.h. $$f(\underline{x} \mid r) = \frac{1}{(2\pi)^{p/2}(\det \boldsymbol{\Sigma}_r)^{1/2}} \cdot \exp\{-\frac{1}{2}(\underline{x} - \underline{\mu}_r)^\top \boldsymbol{\Sigma}_r^{-1}(\underline{x} - \underline{\mu}_r)\}$$

Diskriminanzfunktionen für die Bayes–Zuordnung:

$$\begin{aligned} d_r(\underline{x}) &= \log(f(\underline{x} \mid r)) + \log(p(r)) \\ &= -\frac{1}{2}(\underline{x} - \underline{\mu}_r)^\top \boldsymbol{\Sigma}_r^{-1}(\underline{x} - \underline{\mu}_r) - \frac{1}{2}\log(\det \boldsymbol{\Sigma}_r) + \log(p(r)) \end{aligned}$$

Erster Fall

$$\underline{\underline{X}}|Y = r \quad \sim \quad N(\underline{\mu}_r\,,\ \sigma^2 I) \quad \text{für} \quad r = 1, \ldots, k$$

Erster Fall

Diskriminanzfunktion wird zu: $d_r(\underline{x}) = -\frac{1}{2\sigma^2}||x - \mu_r||^2 + \log(p(r))$

- Bei Vergleich zweier Klassen:

Zweiter Fall

$$\underline{\underline{X}}|Y = r \quad \sim \quad N(\underline{\mu}_r\,,\ \mathbf{\Sigma}) \quad \text{für} \quad r = 1, \ldots, k$$

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

Zweiter Fall

Dritter Fall

$$\underline{\underline{X}}|Y = r \quad \sim \quad N(\underline{\mu}_r\,,\ \boldsymbol{\Sigma}_r) \quad \text{für} \quad r = 1, \ldots, k$$

$$d_r(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_r)^\top \boldsymbol{\Sigma}_r^{-1}(\underline{x} - \underline{\mu}_r) + \log(p(r))$$