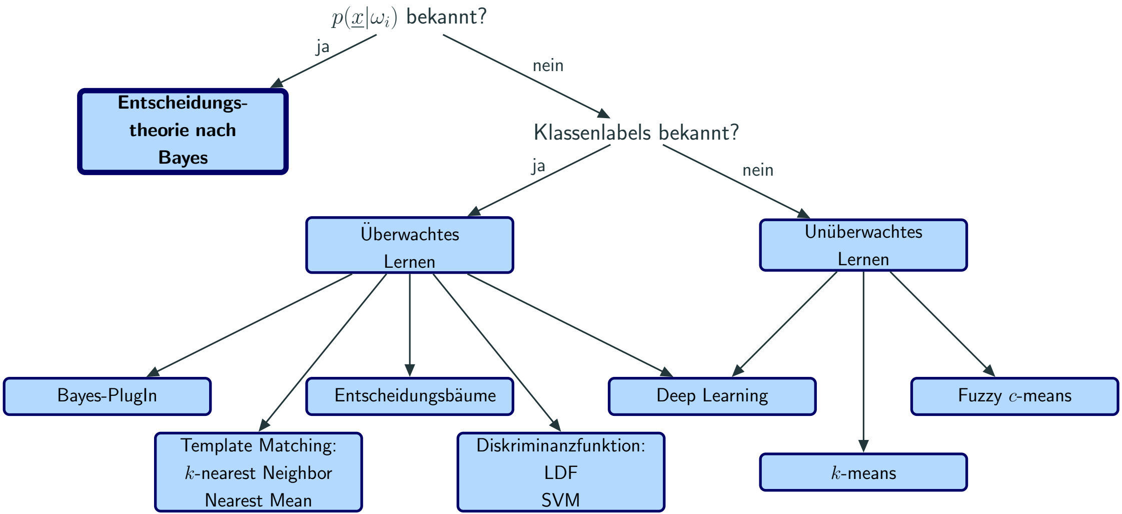
Kapitel 4 - Probabilistische Entscheidungstheorie nach Bayes

Annika Liebgott

November 15, 2020

Überblick



$p(\underline{x}|\omega_i)$ bekannt?
ja
nein
Entscheidungs-theorie nach Bayes
Klassenlabels bekannt?
ja
nein
Überwachtes Lernen
Unüberwachtes Lernen
Bayes-PlugIn
Entscheidungsbäume
Deep Learning
Fuzzy c-means
Template Matching: k-nearest Neighbor Nearest Mean
Diskriminanzfunktion: LDF SVM
k-means

Definitionen + Rückblick Kapitel 2.2 (1)

Definitionen

$\omega \in \{\omega_1, \dots, \omega_c\}$: unbekannte, zu schätzende Klasse (wahre Klasse)

$\underline{x} \in \mathbb{R}^d$: d-dimensionaler Merkmalsvektor

$\hat{\omega}$: geschätzte Klasse

Definitionen

$\omega \in \{\omega_1, \ldots, \omega_c\}$: unbekannte, zu schätzende Klasse (wahre Klasse)

$\underline{x} \in \mathbb{R}^d$: d-dimensionaler Merkmalsvektor

$\hat{\omega}$: geschätzte Klasse

Rückblick: Wahrscheinlichkeiten

$P(\omega = \omega_i) = P(\omega_i) = P_i$: A-Priori-WK von Klasse ω_i, $\quad \sum_{i=1}^{c} P_i = 1$

$P(\omega_i|\underline{x})$: A-Posteriori-WK

$P_{ij} = P(\hat{\omega} = \omega_i, \omega = \omega_j)$: WK, dass ω_i geschätzt wird, während ω_j wahr ist
$\Rightarrow$ Verbundwahrscheinlichkeit, $\quad \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} = 1$

Definitionen + Rückblick Kapitel 2.2 (2)

Rückblick: Verteilungen

$p(\underline{x}|\omega_i)$: klassenabhängige PDF von $\underline{x} \Rightarrow$ "Likelihood" von $\underline{x}$

$p(\underline{x})$: Marginalverteilung von $\underline{x} \Rightarrow$ Evidenz von $\underline{x}$

Definitionen + Rückblick Kapitel 2.2 (2)

Rückblick: Verteilungen

$p(\underline{x}|\omega_i)$: klassenabhängige PDF von $\underline{x} \Rightarrow$ "Likelihood" von $\underline{x}$

$p(\underline{x})$: Marginalverteilung von $\underline{x} \Rightarrow$ Evidenz von $\underline{x}$

Rückblick: Bayes-Theorem

$$P(B|A) = \frac{P(A,B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

4.1 - Minimum Bayesian Risk

Minimum Bayesian Risk (MBR)-Klassifikator

Grundidee: Klasse ω_i, für die Wert von BR minimal wird, wird ausgewählt

Berechnung von $\hat{\omega}$:

$$\hat{\omega}_{\text{BR}} = \arg\min_{\hat{\omega}} \text{BR} = \arg\min_{\hat{\omega}} R(\hat{\omega}|\underline{x}) = \arg\min_{\hat{\omega}} \sum_{j=1}^{c} l(\hat{\omega}(\underline{x}), \omega = \omega_j) \cdot P(\omega = \omega_j|\underline{x})$$

$$f_i(\underline{x}) = -R(\hat{\omega}(\underline{x}) = \omega_i|\underline{x}) \text{ bzw. } f_i(\underline{x}) = \sum_{j=1}^{c} l_{ij} P(\omega_j|\underline{x})$$

Vorgehensweise:

1. Berechne $R(\hat{\omega}(\underline{x}) = \omega_i|\underline{x})$ für alle c möglichen Klassen.
2. Wähle die Klasse, aus der der kleinste Wert für R resultiert.

Maximum A Posteriori (MAP)-Klassifikator

Grundidee: Klasse ω_i, für die A-Posteriori-WK maximal wird, wird ausgewählt.

Berechnung von $\hat{\omega}$:

$$\hat{\omega}_{\text{MAP}} = \arg\max_{\omega_i} P(\omega_i|\underline{x})$$

$$f_i(\underline{x}) = P(\omega_i|\underline{x}) \text{ bzw. } f_i(\underline{x}) = p(\underline{x}|\omega i) \cdot P(\omega_i)$$

Eigenschaften:

- Vorraussetzung: $0/1$-Loss
- $\min$ BR $\hat{=} \max$ A Posteriori

$\Rightarrow$ MAP ist ein Spezialfall von MBR!

Maximum Likelihood (ML)-Klassifikator

Grundidee: Klasse ω_i, für die Likelihood maximal wird, wird ausgewählt.

Berechnung von $\hat{\omega}$:

$$\hat{\omega}_{\mathsf{ML}} = \arg\max_{\omega_i} p(\underline{x}|\omega i)$$

$$f_i(\underline{x}) = p(\underline{x}|\omega_i)$$

Eigenschaften:

- Vorraussetzung 1: $0/1$-Loss
- Voraussetzung 2: gleiche A-Priori-WKs $\Rightarrow P(\omega_j) = \frac{1}{c}$
- $\min$ BR $\hat{=} \max$ A Posteriori $\hat{=} \max$ Likelihood

$\Rightarrow$ ML ist ein Spezialfall von MAP!

Zusammenfassung: MBR, MAP und ML

Zusammenhang:

$$\text{MBR} \xrightarrow{0/1\text{-Loss}} \text{MAP} \xrightarrow{P(\omega_j)=\frac{1}{c}} \text{ML}$$

Voraussetzung:

☑ Likelihood

☑ Loss

☑ A-Priori-Wahrscheinlichkeiten

Aber: Voraussetzungen sind nicht immer bekannt oder variabel.

Frage: Was tun, wenn Voraussetzungen unbekannt oder variabel sind?

A-Priori-Wahrscheinlichkeiten: schwer zu bestimmen?

In manchen Fällen sind A-Priori-WKs einfach zu definieren, weil sie beispielsweise aus empirischen Untersuchungen bekannt oder statistisch berechenbar sind. In anderen Fällen sind sie nicht trivial.

Altersklassifikation: Soll das Bild einer in Deutschland lebenden Person einer Altersgruppe zugeordnet werden, stehen (theoretisch) aus Statistiken der Einwohnermeldeämter A-Priori-WKs zur Verfügung.

Obstsortierung: Je nach Wetterlage ändern sich die Mengenverhältnisse an Obst unterschiedlicher Sorten, die in Deutschland geerntet werden, von Jahr zu Jahr.

Entscheidung in einem Mordfall: Worauf sollen Wahrscheinlichkeiten basieren, ob ein Verdächtiger der Täter ist oder nicht?

Militärische Zielerfassung: Wie legt man Wahrscheinlichkeiten fest, ob eine Person im feindlichen Gebiet Feind oder Zivilist ist?

4.2 - Minimax

Minimax-Klassifikator

Grundidee: Minimierung des maximalen Bayesian Risks

Bekannt: Likelihood, Loss
Unbekannt: A-Priori-Wahrscheinlichkeiten

Vorgehen:

1. Optimiere Klassenwahrscheinlichkeiten $P(\omega_i)$ so, dass MBR maximal wird
2. Wähle anschließend die Klasse $\hat{\omega}$, für die das in Schritt 1 berechnete MBR minimal ist

$\Rightarrow$ Wähle den Fall aus, für den der worst case am wenigsten schlimm ist.

Zusammenfassung Minimax-Klassifikation

Berechnung von $\hat{\omega}$

$$\hat{\omega} = \arg\min_{\omega_i} \max_{P(\omega_i)} \mathrm{BR}$$

Voraussetzung:

- ☑ Likelihood
- ☑ Loss
- ☐ A-Priori-Wahrscheinlichkeiten

4.3 - Neyman-Pearson

Neyman-Pearson-Klassifikator

Grundidee: Klassifikator so designen, dass die False Alarm Rate (P_{FA}) unterhalb eines festgelegten Schwellwerts α liegt. Anderer Name: CFAR (Constant False Alarm Rate)

Bekannt: Likelihood

Unbekannt: A-Priori-Wahrscheinlichkeiten, Loss

2-Klassen-Problem:

$$\begin{bmatrix} 1 - P_{\text{FA}} & 1 - P_{\text{D}} \\ P_{\text{FA}} & P_{\text{D}} \end{bmatrix}$$

Vorgehen:

- maximiere die Detection Rate (P_{D}) so, dass gleichzeitig $P_{\text{FA}} \leq \alpha$ gilt
- Durchführung eines Likelihood-Ratio-Tests mit einem von α abhängigem Schwellwert $\gamma(\alpha)$

Zusammenfassung Neyman-Pearson-Klassifikation

Berechnung von $\hat{\omega}$

$$\hat{\omega} = \arg\max_{\omega_i} P_{\mathsf{D}}$$
$$\text{s.t.} \quad P_{\mathsf{FA}} \leq \alpha$$

Voraussetzung:

- ☑ Likelihood
- ☐ Loss
- ☐ A-Priori-Wahrscheinlichkeiten