

6.6 Schätzung der klassenspezifischen Dichten

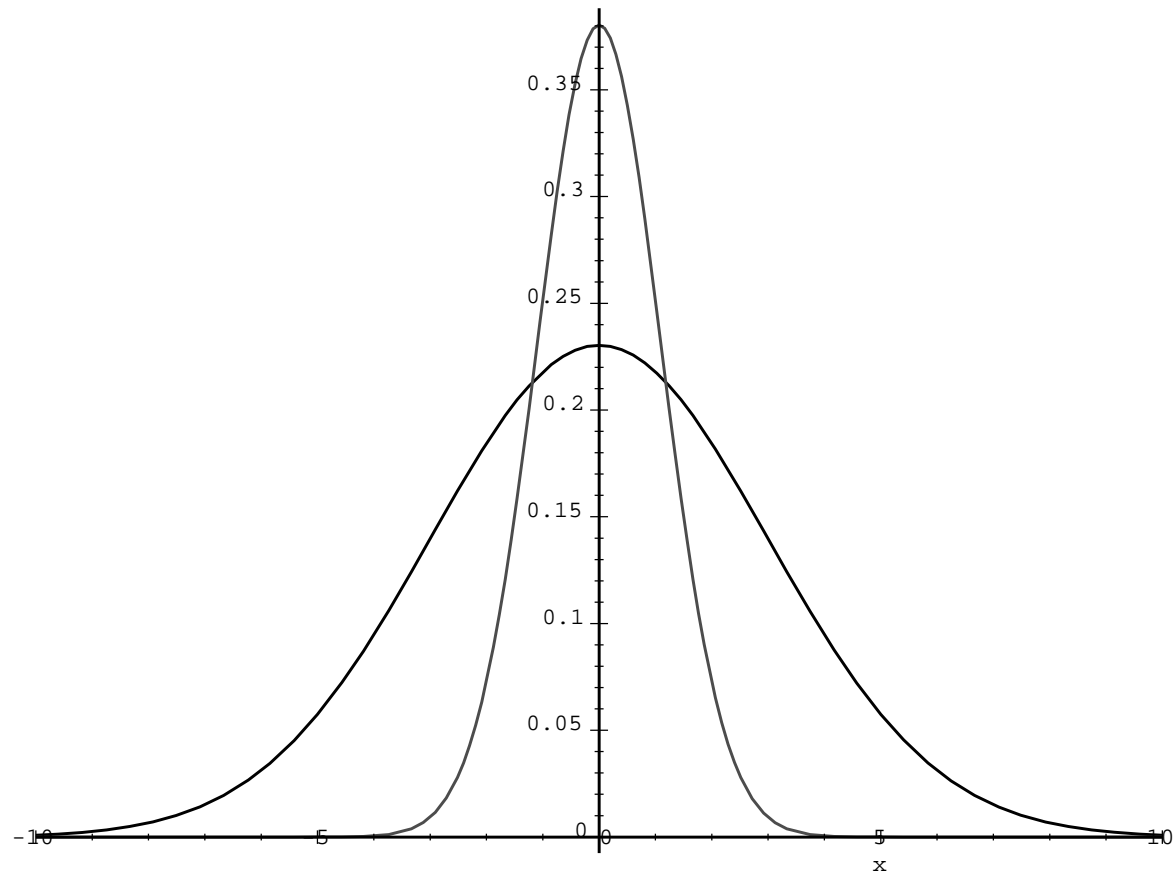
6.6.1 Normalverteilung

Um die Verteilung einer Zufallsvariable X zu modellieren, geht man oft davon aus, daß sie *normalverteilt* ist, d.h.

$$P(X = x) = \mathcal{N}_x(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- $\mu = E\{X\}$ Erwartungswert von X
bestimmt das Zentrum der Normalverteilung
- $\sigma^2 = E\{(X - \mu)^2\}$ Varianz
bestimmt Breite und Höhe der Kurve
- Ist die Näherung durch eine einfache Normalverteilung nicht geeignet, so lassen sich in den allermeisten Fällen durch einfache Überlagerung mehrerer Normalverteilungen gute Ergebnisse erzielen.

6.6 Schätzung der klassenspezifischen Dichten



Beispiel zweier Normalverteilungen im \mathbb{R}^1

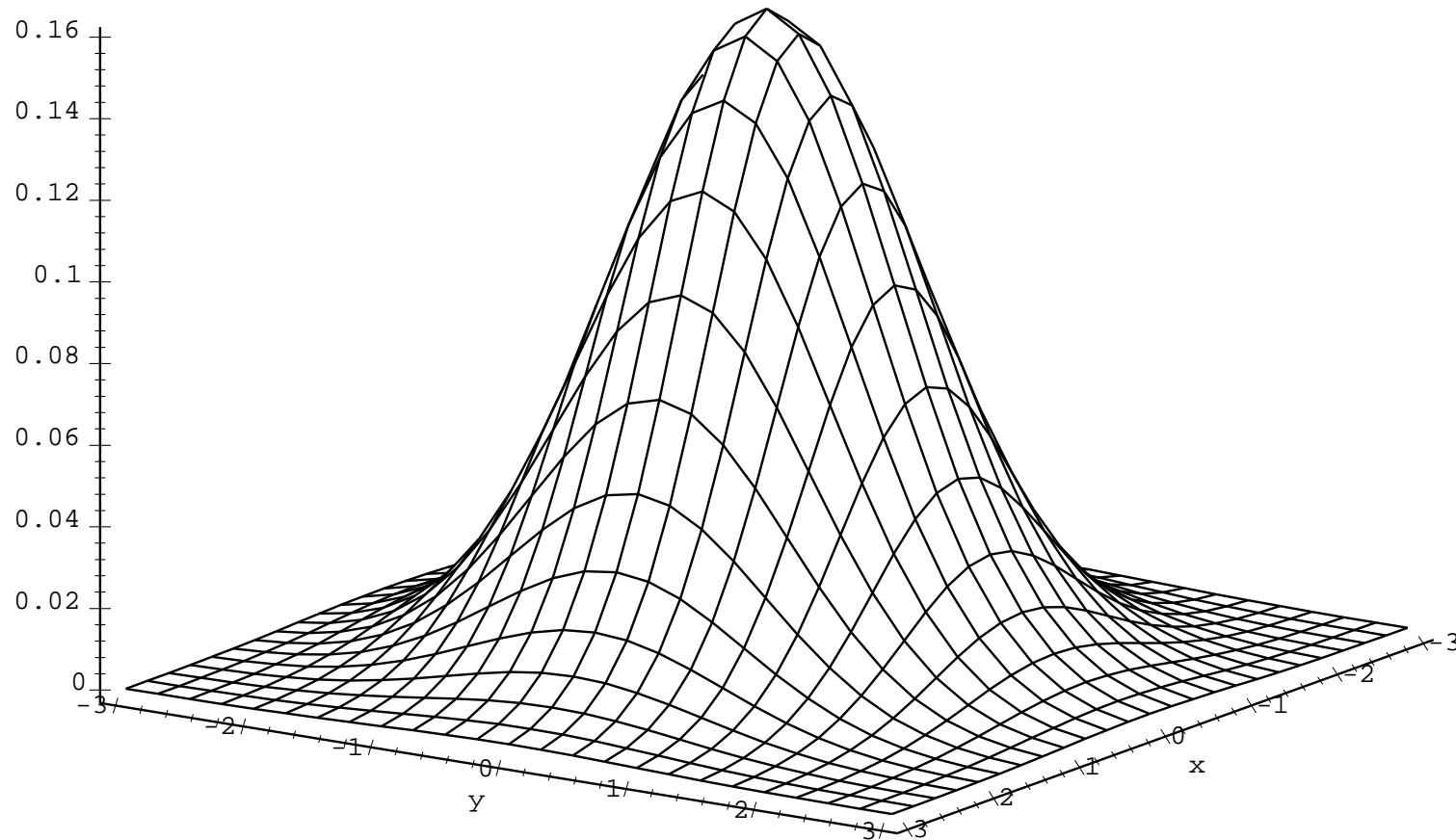
6.6 Schätzung der klassenspezifischen Dichten

Ist die Zufallsvariable ein N -dimensionaler Vektor \vec{c} , so gilt

$$P(\vec{c}) = \mathcal{N}_{\vec{c}}(\vec{\mu}, \underline{K}) = \frac{1}{\sqrt{(2\pi)^N \det \underline{K}}} e^{-\frac{1}{2}(\vec{c}-\vec{\mu})^T \underline{K}^{-1}(\vec{c}-\vec{\mu})}$$

- $\vec{\mu} = E\{\vec{c}\}$ Erwartungswert von \vec{c}
- $\underline{K} = E\left\{(\vec{c} - \vec{\mu})(\vec{c} - \vec{\mu})^T\right\}$ die Kovarianzmatrix.
- Zentrum der Normalverteilung ist — wie im \mathbb{R}^1 — durch den Erwartungswert $\vec{\mu}$ gegeben
- im \mathbb{R}^2 haben Normalverteilungen eine Glockenform
horizontale Schnitte durch diese Glocke sind entweder kreis- oder ellipsenförmig

6.6 Schätzung der klassenspezifischen Dichten



Beispiel einer zweidimensionalen Normalverteilung

6.6.2 Parameterschätzung

schätze Mittelwert $\vec{\mu}$ und Kovarianzmatrix \underline{K} als ML-Schätzwert aus einer Stichprobe der Größe I :

$$\hat{\vec{\mu}} = \frac{1}{I} \sum_{i=1}^I \vec{c}_i$$

$$\hat{\underline{K}} = \frac{1}{I} \sum_{i=1}^I (\vec{c}_i - \hat{\vec{\mu}})(\vec{c}_i - \hat{\vec{\mu}})^T \text{ bzw. } \frac{1}{I-1} \text{ für erwartungstreue Schätzung}$$

Rekursive Berechnung

Es kommt häufig vor, daß die Menge der Trainingsmuster erweitert werden soll:

- neues klassifiziertes Trainingsmaterial vorhanden
- durchführen von entscheidungsüberwachtem Lernen

Da es für umfangreiche Stichproben sehr aufwendig ist, bei jeder Erweiterung die Parameter $\hat{\vec{\mu}}$ und $\hat{\underline{K}}$ vollständig neu zu berechnen, ist eine rekursive Formel für die

6.6 Schätzung der klassenspezifischen Dichten

Schätzwerte von großem Vorteil.

Formel für die rekursive Berechnung des Erwartungswertes:

$$\begin{aligned}\widehat{\vec{\mu}}_I &= \frac{1}{I} \sum_{i=1}^I \vec{c}_i \\ &= \frac{1}{I} \left(\sum_{i=1}^{I-1} \vec{c}_i \right) + \frac{1}{I} \vec{c}_I \\ &= \frac{1(I-1)}{I(I-1)} \left(\sum_{i=1}^{I-1} \vec{c}_i \right) + \frac{1}{I} \vec{c}_I \\ &= \frac{I-1}{I} \cdot \left(\frac{1}{I-1} \sum_{i=1}^{I-1} \vec{c}_i \right) + \frac{1}{I} \vec{c}_I \\ &= \left(1 - \frac{1}{I} \right) \cdot \widehat{\vec{\mu}}_{I-1} + \frac{1}{I} \vec{c}_I\end{aligned}$$

6.6 Schätzung der klassenspezifischen Dichten

Ähnlich läßt sich die Formel für die Kovarianzmatrix herleiten:

$$\underline{\hat{K}}_I = \left(1 - \frac{1}{I}\right) \left[\underline{\hat{K}}_{I-1} + \frac{1}{I} (\vec{c}_I - \hat{\vec{\mu}}_{I-1})(\vec{c}_I - \hat{\vec{\mu}}_{I-1})^T \right]$$

Sogar für die Inverse der Kovarianzmatrix gibt es eine inverse Formel:

$$\underline{\hat{K}}_I^{-1} = \frac{I}{I-1} \left[\underline{\hat{K}}_{I-1}^{-1} - \frac{1}{I} \frac{\underline{\hat{K}}_{I-1}^{-1} (\vec{c}_I - \hat{\vec{\mu}}_{I-1})(\vec{c}_I - \hat{\vec{\mu}}_{I-1})^T \underline{\hat{K}}_{I-1}^{-1}}{1 + \frac{1}{I} (\vec{c}_I - \hat{\vec{\mu}}_{I-1})^T \underline{\hat{K}}_{I-1}^{-1} (\vec{c}_I - \hat{\vec{\mu}}_{I-1})} \right]$$

6.6.3 Bayes-Normalverteilungsklassifikator

Annahme: $P(\vec{c} | \omega_i) = \mathcal{N}_{\vec{c}}(\vec{\mu}_i, \underline{K}_i)$

- schätze klassenspezifischen Mittelwerte $\vec{\mu}_l$ und Kovarianzmatrizen \underline{K}_l aus klassifizierter Stichprobe
für Schätzung von $\vec{\mu}_l$ und \underline{K}_l verwende nur die Merkmalsvektoren \vec{c}_i , die aus der Klasse ω_l stammen
- schätze a priori Wahrscheinlichkeiten als relative Häufigkeit jeder Klasse in der Stichprobe
- für die Unterscheidungsfunktion des Bayes-Klassifikators benötigt man jedoch die Wahrscheinlichkeiten $P(\omega_i | \vec{c})$
gemäß dem Gesetz von Bayes und Randdichten gilt jedoch:

$$P(\omega_i | \vec{c}) = \frac{P(\vec{c} | \omega_i)P(\omega_i)}{P(\vec{c})} = \frac{P(\vec{c} | \omega_i)P(\omega_i)}{\sum_{l=1}^K P(\vec{c}, \omega_l)} = \frac{P(\vec{c} | \omega_i)P(\omega_i)}{\sum_{l=1}^K P(\vec{c} | \omega_l)P(\omega_l)}$$

6.6 Schätzung der klassenspezifischen Dichten

- verwendet man den Bayes-Klassifikator ohne Rückweisung, so reduziert sich der Rechenaufwand deutlich:

$P(\vec{c})$ kann als Konstante vernachlässigt werden, damit ergibt sich:

Die i -te Komponente $d_i(\vec{c})$ der Unterscheidungsfunktion $\vec{d}(\vec{c})$ hat dann die Form

$$d_i(\vec{c}) = P(\vec{c} | \omega_i) \cdot P(\omega_i) = \frac{P(\omega_i)}{\sqrt{(2\pi)^N \det(\underline{K}_i)}} e^{-\frac{1}{2}(\vec{c} - \vec{\mu}_i)^T \underline{K}_i^{-1} (\vec{c} - \vec{\mu}_i)}$$

- die Rangordnung der Ergebnisse der Unterscheidungsfunktion ändert sich nicht (bzw. dreht sich um), wenn
 - man sie mit positiven (negativen) Konstanten multipliziert
 - oder Konstanten addiert
 - oder Logarithmiert

damit:

$$d'_i(\vec{c}) = \underbrace{-2 \ln(P(\omega_i)) + \ln(\det(\underline{K}_i))}_{\text{Konstante } b_{0i}} + \underbrace{(\vec{c} - \vec{\mu}_i)^T \underline{K}_i^{-1} (\vec{c} - \vec{\mu}_i)}_{\text{je Klass. zu berechnen}}$$

6.6 Schätzung der klassenspezifischen Dichten

- Bayesklassifikator für normalverteilte Klassen ist also ein *quadratischer* Klassifikator (Aufwand $O(N^2)$ mit $N = \dim(\vec{c})$)
- mit der Unterscheidungsfunktion

$$\vec{d}(\vec{c}) = \begin{pmatrix} b_{01} + (\vec{c} - \vec{\mu}_1)^T \underline{K}_1^{-1} (\vec{c} - \vec{\mu}_1) \\ \vdots \\ b_{0i} + (\vec{c} - \vec{\mu}_i)^T \underline{K}_i^{-1} (\vec{c} - \vec{\mu}_i) \\ \vdots \\ b_{0K} + (\vec{c} - \vec{\mu}_K)^T \underline{K}_K^{-1} (\vec{c} - \vec{\mu}_K) \end{pmatrix}$$

wird das Risiko mit folgender Entscheidungsregel minimiert:

$$g(\vec{c}) = \hat{\omega} = e(\vec{d}(\vec{c})) = \omega_l, \quad \text{falls } l \text{ minimale Komponente von } \vec{d}(\vec{c})$$

- Berechtigung der Annahme normalverteilter Klassen:
 - statistische Tests
 - Annahme anhand des Klassifikationsergebnisses auf einer Teststichprobe evaluieren
(diese liefert aber in keinem Fall eine Aussage über die Korrektheit der Normalverteilungs-Annahme)

6.6.4 Vereinfachung durch identische Kovarianzmatrizen

Annahme: die Kovarianzmatrizen aller Klassen sind ω_i identisch

- dann vereinfacht sich $d'_i(\vec{c})$ weiter, da $\ln(\det(\underline{K}))$ nun konstant:

$$\begin{aligned}d'_i(\vec{c}) &= -2 \ln(P(\omega_i)) + (\vec{c} - \vec{\mu}_i)^T \underline{K}^{-1} (\vec{c} - \vec{\mu}_i) \\ &= -2 \ln(P(\omega_i)) + \vec{c}^T \underline{K}^{-1} \vec{c} - 2 \vec{\mu}_i^T \underline{K}^{-1} \vec{c} + \vec{\mu}_i^T \underline{K}^{-1} \vec{\mu}_i\end{aligned}$$

- Der Term $\vec{c}^T \underline{K}^{-1} \vec{c}$ kann ebenfalls vernachlässigt werden:

$$d'_i(\vec{c}) = \underbrace{-2 \ln(P(\omega_i)) + \vec{\mu}_i^T \underline{K}^{-1} \vec{\mu}_i}_{b_{1i}} - \underbrace{2 \vec{\mu}_i^T \underline{K}^{-1} \vec{c}}_{b_{2i}}$$

6.6 Schätzung der klassenspezifischen Dichten

- mit der Unterscheidungsfunktion

$$\vec{d}(\vec{c}) = \begin{pmatrix} b_{11} - b_{21}\vec{c} \\ \vdots \\ b_{1i} - b_{2i}\vec{c} \\ \vdots \\ b_{1K} - b_{2K}\vec{c} \end{pmatrix}$$

so wird das Risiko minimiert, falls man folgende Entscheidungsregel anwendet:

$$g(\vec{c}) = \hat{\omega} = e(\vec{d}(\vec{c})) = \omega_l, \quad \text{falls } l \text{ minimale Komponente von } \vec{d}(\vec{c})$$

- Bayesklassifikator für normalverteilte Klassen mit identischer Kovarianzmatrix ist also ein *linearer* Klassifikator (Aufwand $O(N)$ mit $N = \dim(\vec{c})$)
⇒ oft deutliche Rechenzeiterparnis, da Merkmalsvektoren mit 10 bis 100 Dimensionen nicht selten

6.6.5 Klassengrenzen

- ein Merkmalsvektor \vec{c} liegt auf einer Klassengrenze, falls mehrere Komponenten der Unterscheidungsfunktion $\vec{d}(\vec{c})$ den gleichen (maximalen bzw. minimalen) Wert annehmen:

$$\begin{aligned}d_{\kappa}(\vec{c}) &= d_{\lambda}(\vec{c}) \Leftrightarrow \\d_{\kappa}(\vec{c}) - d_{\lambda}(\vec{c}) &= 0\end{aligned}$$

- Einsetzen für Bayesklassifikators mit Normalverteilungen ergibt:

$$d_{\kappa}(\vec{c}) - d_{\lambda}(\vec{c}) = b_{0\kappa} + (\vec{c} - \vec{\mu}_{\kappa})^T \underline{K}_{\kappa}^{-1} (\vec{c} - \vec{\mu}_{\kappa}) - b_{0\lambda} - (\vec{c} - \vec{\mu}_{\lambda})^T \underline{K}_{\lambda}^{-1} (\vec{c} - \vec{\mu}_{\lambda}) = 0$$

- läßt sich durch geeignete Transformationen zu einem einzigen quadratischen Ausdruck umformen
 \Rightarrow alle Klassengrenzen sind Ellipsen (Ellipsoide), Parabel (Paraboloide) oder Hyperbeln (Hyperboloide)

6.6 Schätzung der klassenspezifischen Dichten

- Bayesklassifikator für Normalverteilungen mit identischen Kovarianzmatrizen:

$$d_{\kappa}(\vec{c}) - d_{\lambda}(\vec{c}) = b_{1\kappa} - b_{2\kappa}\vec{c} - b_{1\lambda} + b_{2\lambda}\vec{c}$$

diese Gleichung ist linear

⇒ Klassengrenzen sind durch Geraden, Flächen oder Hyperflächen gegeben

6.6.6 Mischverteilungen

- Verteilung wird durch Linearkombination von verschiedenen (Normal)verteilungen beschrieben:

$$P(\vec{c} \mid \omega_i) = \sum_{l=1}^L a_l \cdot \mathcal{N}_{\vec{c}}(\vec{\mu}_l, \underline{K}_l)$$

wobei $\vec{\theta} = (a_1, \vec{\mu}_1, \underline{K}_1, \dots, a_L, \vec{\mu}_L, \underline{K}_L)$

die Mischverteilung (hier der Klasse ω_i) beschreibt

- Schätzen im Prinzip wie k-means, wobei nun jedes Gebiet R_l bzw. jede l-te Normalverteilung (nicht **Klasse**)
 - nicht mehr alleine durch $\vec{\mu}_l$, sondern
 - durch Parameter $\vec{\mu}_l$ und \underline{K}_l einer hochdimensionalen Normalverteilung $\mathcal{N}_{\vec{c}}(\vec{\mu}_l, \underline{K}_l)$ bestimmt ist,
 - $P(\vec{c} \mid R_l) = \mathcal{N}_{\vec{c}}(\vec{\mu}_l, \underline{K}_l)$
 - $P(R_l) = a_l$

Harte Vektorquantisierung

- ordne innerhalb jeder Iteration jeden Merkmalsvektor \vec{c}_i einem Gebiet **hart** zu
- Ziel: durchschnittlichen Likelihood-Wert H maximieren bezüglich der Parameter $\vec{\theta}$
(In um Konvergenz zu beweisen)

$$H = \frac{1}{I} \sum_{i=1}^I \ln \max_{R_l} P(\vec{c}_i, R_l | \vec{\theta}) = \frac{1}{I} \sum_{i=1}^I \ln \max_{R_l} \ln(P(R_l)P(\vec{c}_i | R_l))H$$

- Verfahren zur Optimierung von L Gebieten

6.6 Schätzung der klassenspezifischen Dichten

wähle aufgrund von Vorwissen oder zufällig initiale Parameter $P(R_l), \vec{\mu}_l, \underline{K}_l$ (z.B. $P(R_l) = 1/L, \vec{\mu}_l = l$ -ter Vektor der Stichprobe, $\underline{K}_l = \mathcal{I}$)
$H^0 := -\infty$ (Wert der Likelihood-Funktion in der Iteration 0)
$t := 0$ (Iterationszähler)
$t := t + 1, \quad H^{(t)} := 0$
FOR alle Gebiete $R_l, l = 1, \dots, L$
$I_l := 0; \quad \hat{\vec{\mu}}_l := \vec{0}, \quad \hat{\underline{M}}_l := O$
FOR alle Vektoren \vec{c}_i der Stichprobe
bestimme R_l mit maximalem $P(R_l \vec{c}_i)$
$H^{(t)} := H^{(t)} + \ln(P(R_l) P(\vec{c} R_l))$
berechne neue Schätzwerte für den Mittelwert und die Momentenmatrix, d.h.
$\hat{\vec{\mu}}_l := \hat{\vec{\mu}}_l + \vec{c}_i \quad \hat{\underline{M}}_l := \hat{\underline{M}}_l + \vec{c}_i \vec{c}_i^T$
$I_l := I_l + 1$
$H^{(t)} := H^{(t)} / I$
FOR alle Gebiete $R_l, l = 1, \dots, L$
$P(R_l) = \frac{I_l}{I}, \quad \vec{\mu}_l := \hat{\vec{\mu}}_l / I_l; \quad \underline{K}_l := \hat{\underline{M}}_l / I_l - \vec{\mu}_l \vec{\mu}_l^T$
UNTIL $(H^{(t)} - H^{(t-1)}) / H^{(t)} \leq \varepsilon$

Weiche Vektorquantisierung, EM-Algorithmus

- ordne jeden Merkmalsvektor \vec{c}_i mit dem Gewicht seiner a-posteriori-Wahrscheinlichkeit $P(R_l | \vec{c}_i)$ **allen** Gebieten R_l **weich** zu

(beachte: $\sum_{l=1}^L P(R_l | \vec{c}_i) = 1$)

- Ziel: Maximierung (der logarithmierten) Wahrscheinlichkeit H der Produktion der Stichprobe in Abhängigkeit der Mischverteilung (normiert bzgl. der Stichprobengröße I):

$$H = \frac{1}{I} \ln P(\{\vec{c}_1, \dots, \vec{c}_I\} | \vec{\theta}) = \frac{1}{I} \ln \prod_{i=1}^I P(\vec{c}_i | \vec{\theta}) = \frac{1}{I} \sum_{i=1}^I \ln P(\vec{c}_i | \vec{\theta}) =$$
$$\frac{1}{I} \sum_{i=1}^I \ln \sum_{l=1}^L P(R_l) p(\vec{c}_i | R_l)$$

6.6 Schätzung der klassenspezifischen Dichten

- zur Berechnung von $P(R_l | \vec{c}_i)$ folgende Umformungen (Bayes-Gesetzes):

$$\begin{aligned} P(R_l | \vec{c}_i) &= \frac{P(R_l)P(\vec{c}_i | R_l)}{P(\vec{c}_i)} \\ P(\vec{c}_i) &= \sum_{k=1}^L P(\vec{c}_i, R_k) \\ &= \sum_{k=1}^L P(R_k)P(\vec{c}_i | R_k) \\ \Rightarrow P(R_l | \vec{c}_i) &= \frac{P(R_l)P(\vec{c}_i | R_l)}{\sum_{k=1}^L P(R_k)P(\vec{c}_i | R_k)} \end{aligned}$$

- Verfahren zur Optimierung von L Gebieten

6.6 Schätzung der klassenspezifischen Dichten

wähle aufgrund von Vorwissen oder zufällig initiale Parameter $P(R_l), \vec{\mu}_l, \underline{K}_l$ (z.B. $P(R_l) = 1/L, \vec{\mu}_l = l$ -ter Vektor der Stichprobe, $\underline{K}_l = \mathcal{I}$)
$H^0 := -\infty$ (Wert der Likelihood-Funktion in der Iteration 0)
$t := 0$ (Iterationszähler)
$t := t + 1, \quad H^{(t)} := 0$
FOR alle Gebiete $R_l, l = 1, \dots, L$
$I_l := 0; \quad \hat{\vec{\mu}}_l := \vec{0}, \quad \hat{\underline{M}}_l := O$
FOR alle Vektoren \vec{c}_i der Stichprobe
berechne $P(R_l \vec{c}_i)$ mittels $P(\vec{c} R_l) = \mathcal{N}_{\vec{c}}(\vec{\mu}_l, \underline{K}_l)$ und $P(R_l)$ für alle Gebiete R_l
$H^{(t)} := H^{(t)} + \ln\left(\sum_{l=1}^L P(R_l) P(\vec{c} R_l)\right)$
FOR alle Gebiete $R_l, l = 1, \dots, L$
berechne neue Schätzwerte für den Mittelwert und die Momentenmatrix, d.h.
$\hat{\vec{\mu}}_l := \hat{\vec{\mu}}_l + \vec{c}_i \cdot P(R_l \vec{c}_i) \quad \hat{\underline{M}}_l := \hat{\underline{M}}_l + \vec{c}_i \vec{c}_i^T \cdot P(R_l \vec{c}_i)$
$I_l := I_l + P(R_l \vec{c}_i)$
$H^{(t)} := H^{(t)} / I$
FOR alle Gebiete $R_l, l = 1, \dots, L$
$P(R_l) = \frac{I_l}{I}, \quad \vec{\mu}_l := \hat{\vec{\mu}}_l / I_l; \quad \underline{K}_l := \hat{\underline{M}}_l / I_l - \vec{\mu}_l \vec{\mu}_l^T$
UNTIL $(H^{(t)} - H^{(t-1)}) / H^{(t)} \leq \varepsilon$

Klassenabhängige Dichten

- zerlege **klassifizierte** Stichprobe in K Stichproben S_k , $k = 1, \dots, K$, die jeweils nur Vektoren aus der Klasse ω_k enthalten
- wende gesondert auf jede dieser Stichproben die Vector Quantization an. Man erhält L_k Dichten mit den Parametern $\mathcal{N}(\vec{\mu}_l^k, \underline{K}_l^k)$
- bestimme aus der klassifizierten Stichprobe die Gewichte wie folgt:

$$a_l^k = \frac{1}{I^k} \sum_{\vec{c}_i \in \omega_k \wedge \vec{c}_i \in R_l^k} 1 \quad (\text{harte VQ})$$

$$a_l^k = \frac{1}{I^k} \sum_{\vec{c}_i \in \omega_k} P(R_l^k \mid \vec{c}_i) \quad (\text{weiche VQ})$$

- die Komponenten der Unterscheidungsfunktion ergeben sich nun zu:

$$P(\vec{c} \mid \omega_k) = \sum_{l=1}^{L_k} a_l^k \cdot \mathcal{N}(\vec{\mu}_l^k, \underline{K}_l^k)$$

Klassenunabhängige Dichten

- schätze auf einer **unklassifizierten** Stichprobe (weiche oder harte VQ)
⇒ L Normalverteilungen $\mathcal{N}(\vec{\mu}_l, \underline{K}_l)$
- bestimme die Gewichtsparemeter der obigen Dichten für die Klasse ω_k aus einer (evtl. kleineren) **klassifizierten** Stichprobe wie folgt:

$$a_l^k = \frac{1}{I^k} \sum_{\vec{c}_i \in \omega_k \wedge \vec{c}_i \in R_l} 1 \quad (\text{harte VQ})$$

$$a_l^k = \frac{1}{I^k} \sum_{\vec{c}_i \in \omega_k} P(R_l \mid \vec{c}_i) \quad (\text{weiche VQ})$$

•

$$P(\vec{c} \mid \omega_k) = \sum_{l=1}^L a_l^k \mathcal{N}(\vec{\mu}_l, \underline{K}_l)$$

- dieses Vorgehen ist u.U. günstiger, das sich mehrere Klassen eine Normalverteilung “teilen” können

6.6.7 Andere Verteilungen

Statistische Unabhängigkeit

- **Annahme** statistischer Unabhängigkeit der Merkmale
(die allerdings in der Regel nicht, höchstens approximativ gegeben ist)

-

$$P(\vec{c} | \omega_k) = \prod_{\nu=1}^N P(c_\nu | \omega_k)$$

- die $P(c_\nu | \omega_k)$ können
 - mit eindimensionalen parametrischen Dichte geschätzt werden
(größere Auswahl als im höherdimensionalen Fall)
 - nach Diskretisierung der Werte c_n :
durch Histogramm als relative Häufigkeiten tabelliert werden

Parzenschätzung \sim radiale Basisfunktionen

- **Motivation:** für höhere Dimensionen ist Diskretisierung des Merkmalsraums bei relativ wenigen Stichprobenelementen schwierig
- \Rightarrow “verschmiere” jeden Beitrag der Stichprobe um seine Position
- Approximation der Dichte durch Überlagerung von Normalverteilungen, (oder auch andere Fensterfunktionen, z.B. Rechteck)

$$P(\vec{c} \mid \omega_k) = \frac{1}{I} \sum_{i=1}^I \mathcal{N}(\vec{c}_i, K)$$

wobei $K = \sigma I$, und σ die Breite des Fensters (des “Verschmierens”) bestimmt

6.7 Alternative Klassifikationsergebnisse

bis jetzt: genau **eine** Klasse oder Rückweisung

Erweiterung weniger endgültige Entscheidung des Klassifikators,
um nachfolgenden Prozesss die Entscheidung zu überlassen
(wobei dieser weitere Evidenzen eingehen lassen kann)
⇒ *principle of least commitment*

Idee

- ordne die Klassen ω_i gemäß abfallendem $P(\omega_i | \vec{c})$
- wähle **Konfidenzschwelle** $\theta \in [\frac{1}{K}; 1]$
- Ergebniss der Klassifikation sind die ersten Klassen, sodass bei minimaler Anzahl an Klassen gilt:

$$\sum_i P(\omega_i | \vec{c}) \geq \theta$$

6.7 Alternative Klassifikationsergebnisse

Bemerkungen durch die Wahl von θ kann die Anzahl an Alternativen gewählt werden:

- $\theta = \frac{1}{K}$: nur beste Klasse, wie bisher
- $\theta = 1$: alle Klassen