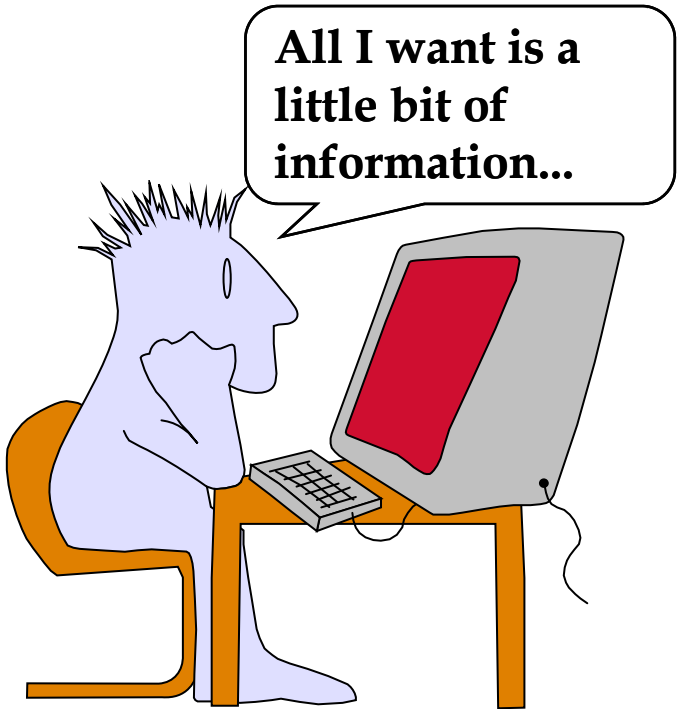
Sprachverarbeitung – 4.Woche

Andreas Wendemuth

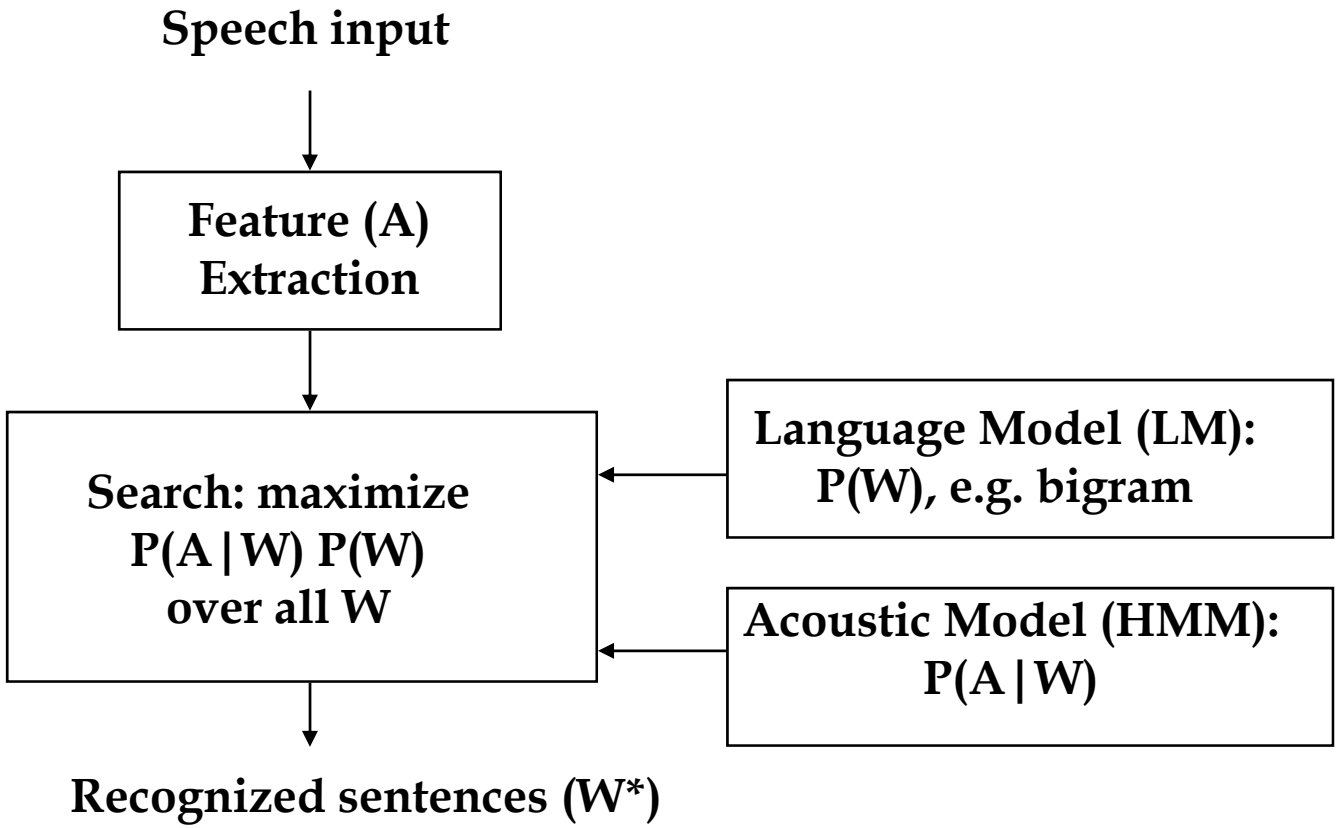


Letzte Woche:

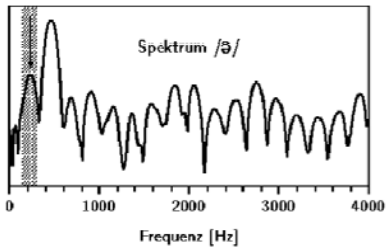
1.Overview Speech Recognition Systems &Architectures
2. Acoustic modeling & feature extraction (1)
3. Feature extraction (2)
4. Klassifikation in HMM Modellen
5. Wortmodellierung (trigramme, tying)
6. search/decoding, lattices,
   wordgraphs, confidence measures
7. acoustic adaptation
8. language models and grammars, Language model
   adaptation, lexica, phonology
9. speech understanding, dialogue control
10. Design of computer speech recognition systems

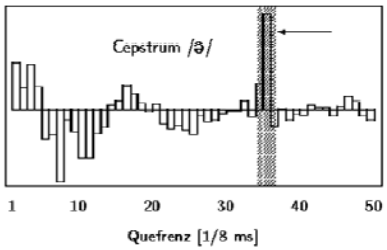
SR Architecture revisited

W = a word sequence (e.g. word/ sentence/ whole dictation)
A = an acoustic feature vector sequence (the input for the recognizer)

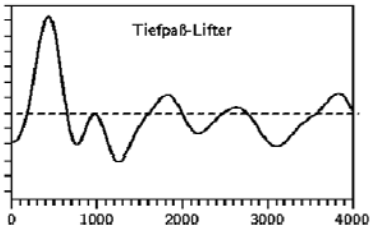


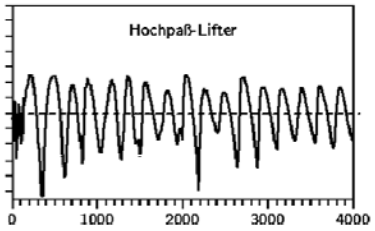
Homomorphe Analyse (Cepstrum)





- Cepstrum ⇔ Spektrum des logarith. Betragsspektrums
- ***Quefrenz*** = Periodendauer in Sekunden (Einheit ist $^1/_{f_A}$)
- ***Cepstralgipfel*** markiert die Dauer der Grundperiode
- ***Formantstruktur*** = niedrige Cepstralkoeffizienten





- ***Lifterung*** — Rücktransformation in den Spektralbereich

$$\{\hat{C}_\nu^{(m)}\} \;=\; \mathrm{DFT}\{\hat{c}_q^{(m)}\}$$

- Grobstruktur: $\hat{c}_q^{(m)} = 0$ für $q > 20$ (Tiefpaß)
- Feinstruktur: $\hat{c}_q^{(m)} = 0$ für $q < 20$ (Hochpaß)

Lineare Vorhersageformel, Ordnung p

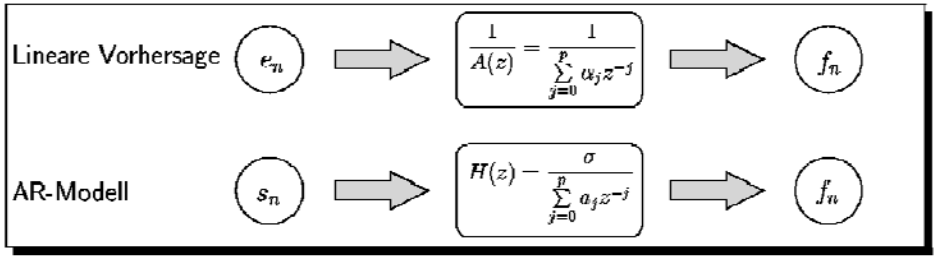
$$\hat{f}_n = -\sum_{j=1}^{p} \alpha_j f_{n-j}$$

Vorhersagefehler mit den Prädiktionskoeffizienten α_j

$$e_n = f_n - \hat{f}_n = \sum_{j=0}^{p} \alpha_j f_{n-j} \qquad (\alpha_0 = 1)$$

Nach z-Transformation ergeben sich die Systemgleichungen

$$E(z) = F(z) \cdot A(z) \quad \text{und} \quad F(z) = E(z) \cdot \frac{1}{A(z)} ,$$



Wenn die Vorhersagekoeffizienten α_j mit den Modellparametern a_j übereinstimmen:

$$e_n = \sigma s_n \qquad \text{(Fehler = Verstärkung} \times \text{Anregung)}$$

Das Modellspektrum

Verknüpfen von Vorhersage- und Produktionsmodell durch die idealisierte Annahme

$$H(z) \;=\; \sigma/A(z)$$

Querschnittflächen A_j aus partieller Korrelation (PARCOR) k_n:

$$A_{j+1} \;=\; \frac{1-k_j}{1+k_j}\cdot A_j$$

Schätzung des **Vokaltraktfrequenzgangs**

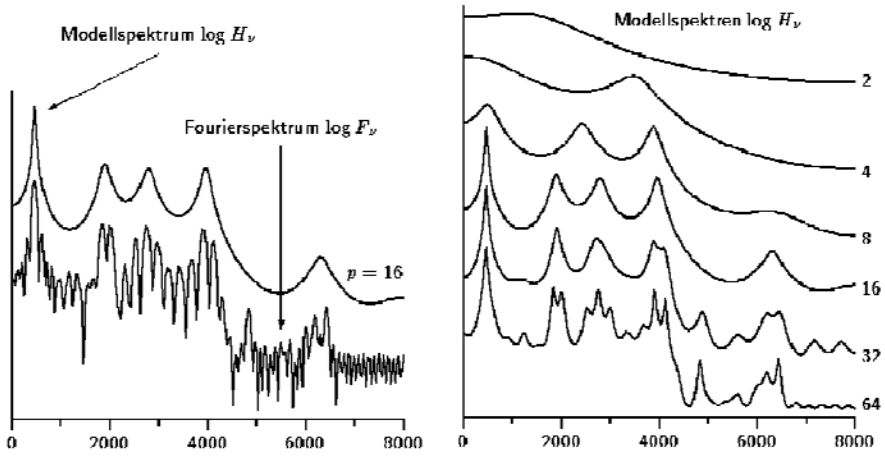
$$H_\nu \;-\; H(e^{2\pi i\nu/N}) \;-\; \frac{\sigma}{A(e^{2\pi i\nu/N}} \;-\; \frac{\sigma}{A_\nu}$$

Funktionswerte A_ν von $A(z)$ auf dem Einheitskreis durch

$$\mathrm{DFT}\{1, \alpha_1, \ldots, \alpha_p, \underbrace{0, \ldots, 0}_{N-p-1 \text{ Nullen}}\}$$

Verstärkungsfaktor aus Autokorrelationsfunktion

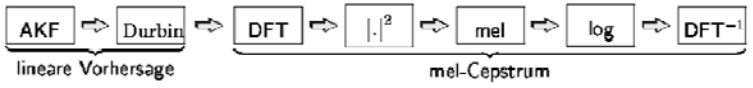
$$\sigma^2 \;-\; \sum_{j=0}^{p} \alpha_j r_j$$



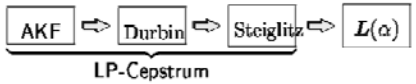
Gehörrichtige Frequenzverzerrung

Integration der Mel-Skalierung in die Berechnung der Vorhersagemerkmale

LP-mel-Cepstrum:

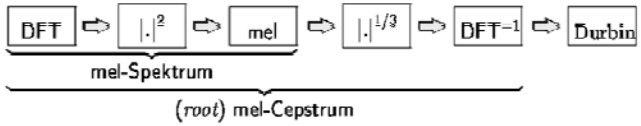


Bilineartransformiertes LP-Cepstrum:



Lineare Vorhersage:

Perzeptuelle lineare Vorhersage (PLP):



Bemerkung 3.4

- $L(\alpha)$ ist eine geeignete ***Bilineartransformation*** zur Frequenzverzerrung im Cepstralbereich
- PLP nach dem Satz von Wiener & Khintchine:

$$\mathrm{FT}\{r_n\} = |\mathrm{FT}\{f_n\}|^2 \quad \text{bzw.} \quad \{r_n\} = \mathrm{FT}^{-1}\{|\mathrm{FT}\{f_n\}|^2\}$$

Diese Woche:

1.Overview Speech Recognition Systems &Architectures
2. Acoustic modeling & feature extraction (1)
3. Feature extraction (2)
4. LDA, Grundlagen der Klassifikation
5. Klassifikation in HMM Modellen
6. Wortmodellierung (trigramme, tying)
7. search/decoding, lattices,
   wordgraphs, confidence measures
8. acoustic adaptation
9. language models and grammars, Language model
   adaptation, lexica, phonology
10. speech understanding, dialogue control
11. Design of computer speech recognition systems

Merkmaltransformationen zur Dimensionsreduktion

- **Suchraum kann massiv eingeschränkt werden durch Dimensionsreduktion**
- **Rauschen kann in „subspaces“ angenommen werden, die durch Dimensionsreduktion verschwinden können**
- **Übersicht:**
  1. Karhunen-Loeve-Transformation (Hauptachsen)
  2. (Quotient) Singular Value Decomposition (Q)SVD
  „Subspace Identification for linear Systems“,
  z.B. Buch (1996): von Overschee and De Moor
  3. Lineare Diskriminanten Analyse (LDA)

Hauptachsentransformation

- - bestimme Drehung des Raumes (Unitäre Transf.)
- - Funktion darstellbar als Produkt in neuen Koordin.
- - Verzerrungsfaktoren sind die Eigenwerte (EW)
- - z.B. schätze die Matrix einer Normalverteilung als Korrelationsmatrix der Beobachtungswerte, auf Hauptachsen stellt die Korr.matrix Ellipsen dar
- - behalte nur die Koordinaten mit den grössten EW

- Problem : - nicht o.k. bei Adidas-Verteilung(s.LDA)
  - berücksichtigt keine Klassenzugehörigkeit

Karhunen-Loéve-Transformation

Un/vollständige Entwicklung von $\boldsymbol{x}$ nach der Basis $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_D$:

$$\boldsymbol{x} = \sum_{i=1}^{D} y_i \boldsymbol{\phi}_i \quad \text{bzw.} \quad \tilde{\boldsymbol{x}} = \sum_{i=1}^{d} y_i \boldsymbol{\phi}_i \quad \text{mit Koeffizienten } y_i = \boldsymbol{\phi}_i^\top \boldsymbol{x}$$

Mittlerer quadratischer Fehler (Dimension d)

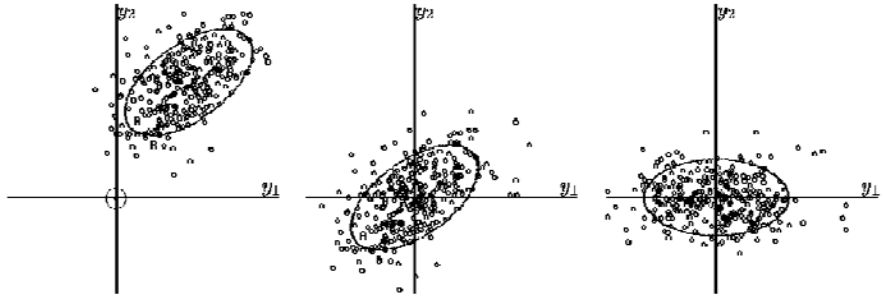
$$\varepsilon_d = \mathcal{E}[\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|^2] = \mathcal{E}[\|\sum_{i=d+1}^{D} y_i \boldsymbol{\phi}_i\|^2] = \sum_{i=d+1}^{D} \boldsymbol{\phi}_i^\top \underbrace{\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\top]}_{:=\boldsymbol{S}} \boldsymbol{\phi}_i$$

Schätzung der **Momentenmatrix** $\boldsymbol{S}$ (Lernstichprobe ω):

$$\hat{\boldsymbol{S}} = \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{x}_j \boldsymbol{x}_j^\top, \quad \omega = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \quad (\boldsymbol{S} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top)$$

Fehlerminimierung $\rightsquigarrow$ **Eigenwertaufgabe** $\boldsymbol{S}\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i$

$$\boldsymbol{S} = \boldsymbol{\Phi}^\top \boldsymbol{\Lambda} \boldsymbol{\Phi}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_D \end{pmatrix}, \quad \lambda_i > 0$$



Die zentrierte Hauptachsen- / KL-transformation

- ... besitzt den Rekonstruktionsfehler $\varepsilon_d = \sum_{i=d+1}^{D} \lambda_i$
- ... hat die Eingabe dekorrelliert ($\boldsymbol{y}$ besitzt Kovarianzmatrix $\boldsymbol{\Lambda}$)

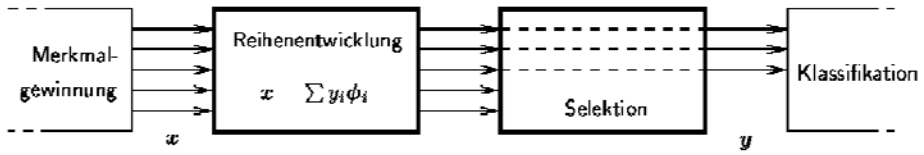
Merkmalraumtransformationen

- **Reduktion der Vektordimension**
  - ⇨ Geringerer Klassifikationsaufwand
  - ⇨ Robustere Schätzung statistischer Parameter
- *(.. möglichst geringer Informationsverlust ...)*

Lineare Transformationsvorschrift

$$\varphi : \begin{cases} \mathbb{R}^D & \rightarrow & \mathbb{R}^d \\ \boldsymbol{x} & \mapsto & \boldsymbol{y} = \varphi(\boldsymbol{x}) = \boldsymbol{\Phi}^\top \boldsymbol{x} \end{cases} \qquad \text{mit } d \leq D$$

Typischerweise ist $\boldsymbol{\Phi}$ eine ***Rotationsmatrix***, d.h. $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \boldsymbol{E}$



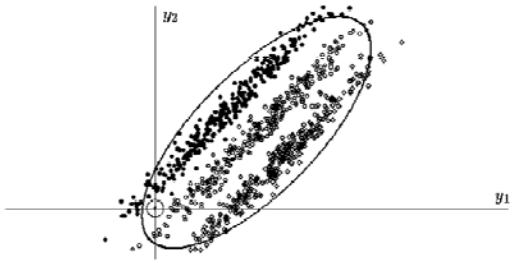
Gütekriterien für $\varphi(\cdot)$

- Varianzmaximierung ⇝ ***Karhunen-Loève-Transformation***
- Klassenseparation ⇝ ***Lineare Diskriminanzanalyse***
- ML-Kriterium ⇝ ***Maximum-Likelihood-Rotationen***
- Fehlerrate ⇝ ***???***

Klassenbezogene Transformation

LINEARE DISKRIMINANZANALYSE

- die **Ballungsgebiete kompakt** halten
- ihre **Zentren** voneinander **entfernen**
- in eine der **Klassentrennung** förderliche Lage **drehen**



Zerlegung der Gesamtkovarianz $\boldsymbol{\Sigma} = \boldsymbol{S}_w + \boldsymbol{S}_b$

$$\boldsymbol{S}_w = \sum_{\kappa=1}^{K} p_\kappa \boldsymbol{\Sigma}_\kappa \qquad \text{(Intraklassen-Streuungsmatrix)}$$

$$\boldsymbol{S}_b = \sum_{\kappa=1}^{K} p_\kappa (\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})(\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})^\top \qquad \text{(Interklassen-Streuungsmatrix)}$$

Zentrierte LDA-Transformation

$$\boldsymbol{y} = \boldsymbol{\Phi}^\top (\boldsymbol{x} - \boldsymbol{\mu}) \quad \text{mit} \quad \boldsymbol{S}_w^{-1} \boldsymbol{S}_b \boldsymbol{\Phi} = \boldsymbol{\Phi} \boldsymbol{\Lambda}$$

Translation — Drehung — Reskalierung

Gauß-Quantisierung des LDA-Cepstrums

Merkmalberechnung

Z.B. die Cepstralkoeffizienten & Ableitungen

Beschaffung einer Laut-Etikettierung

Per „bootstrap“ aus einem initialen Erkennungssystem

Berechnung der LDA-Transformationsmatrix Φ

Eigenwertanalyse der zerlegten Gesamtstreuung

Transformation der Lernstichprobe

Alle $\boldsymbol{x} \in \omega$ werden gemäß $\boldsymbol{\Phi} : \mathbb{R}^{72} \rightarrow \mathbb{R}^{20}$ reduziert

Statistischer Vektorquantisierer-Entwurf

EM-Algorithmus $\rightsquigarrow$ Kodebuch ($K = 256$) für $\omega^{\boldsymbol{\Phi}}$

Quantisierung der Merkmalvektoren

Abbildung der $\boldsymbol{y} \in \mathbb{R}^{20}$ auf 8-bit-Klassenindizes

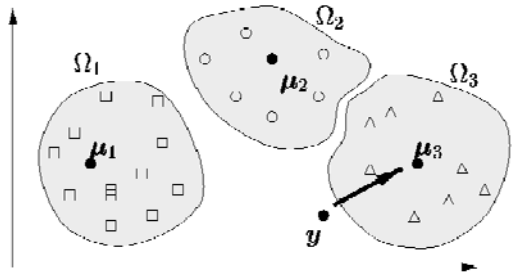
Klassifikation

Klassifikation D-dimensionaler Muster

- Merkmalvektoren $\boldsymbol{x} \in \mathbb{R}^D$
- Musterklassen $\Omega_1, \ldots, \Omega_K$
- Etikettierte Lernstichprobe $\omega_1, \ldots, \omega_K$
- Klassifikationsaufgabe

$$\boldsymbol{y} \in \mathbb{R}^D \quad \rightsquigarrow \quad \Omega_k \qquad (k = ?)$$

⇨ *ÜBERWACHTES LERNEN*

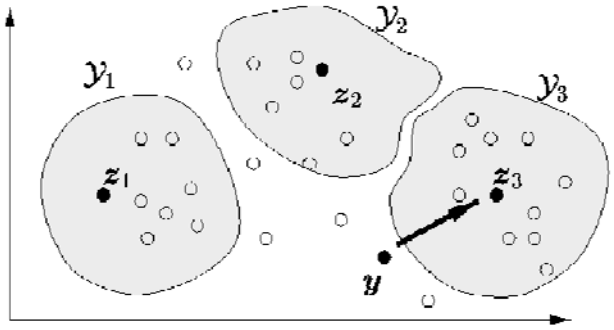


Vektorquantisierung

- Merkmalvektoren $\boldsymbol{x} \in \mathbb{R}^D$
- Repräsentantenvektoren $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$
- Nichtetikettierte Lernstichprobe $\omega \subseteq \mathbb{R}^D$
- Quantisierungsaufgabe

$$\boldsymbol{y} \in \mathbb{R}^D \quad \rightsquigarrow \quad \boldsymbol{z}_k \quad (k = ?)$$

⇨ *UNÜBERWACHTES LERNEN*



Numerische Klassifikatoren

GEGEBEN:

- D-dimensionale, reelle Merkmalvektoren

$$\boldsymbol{x} = (x_1, \ldots, x_D) \in \mathbb{R}^D$$

- Inventar von K Musterklassen

$$\Omega_1, \Omega_2, \ldots, \Omega_K$$

Statistisches Ereugungsmodell:

- Diskreter Zufallsprozeß
  mit a priori Klassenwahrscheinlichkeiten

$$P(\Omega_\kappa)\ , \quad \kappa = 1, \ldots, K \quad \text{mit} \quad \sum_{\kappa=1}^{K} P(\Omega_\kappa) = 1$$

- Multivariat-kontinuierlicher Prozeß
  mit den bedingten Wahrscheinlichkeitsdichtefunktionen

$$P(\boldsymbol{x} \mid \Omega_\kappa) \quad \text{mit} \quad \int_{\mathbb{R}^D} P(\boldsymbol{y} \mid \Omega_\kappa) d\boldsymbol{y} = 1$$

Optimale Entscheidungsregel

„Unscharfe" Entscheidungsregel:

$$\delta(\Omega_\kappa \mid \boldsymbol{x}) \quad \text{mit} \quad \sum_{\kappa=1}^{K} \delta(\Omega_\kappa \mid \boldsymbol{x}) = 1 \quad \text{für alle } \boldsymbol{x} \in \mathbb{R}^D$$

Risikofunktion:

$$R(\delta) = \sum_{\kappa=1}^{K} P(\Omega_\kappa) \sum_{\lambda=1}^{K} r_{\kappa\lambda} \int_{\mathbb{R}^D} \delta(\Omega_\lambda \mid \boldsymbol{x}) P(\boldsymbol{x} \mid \Omega_\kappa) d\boldsymbol{x}$$

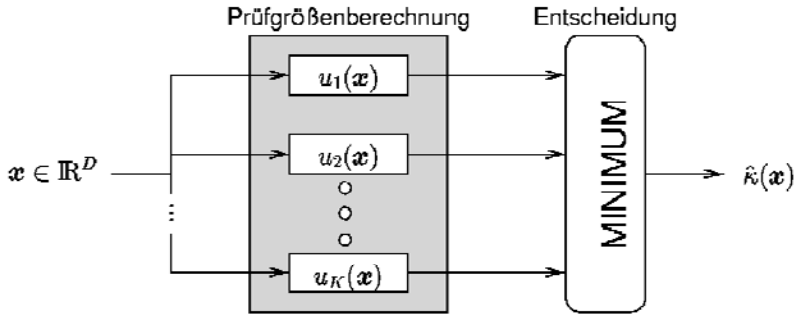
... mit den ***Kosten*** $r_{\kappa\lambda}$ (Verlust, Schaden) für die Verwechslung der Klassen $\Omega_\kappa, \Omega_\lambda$

Die **optimale Entscheidungsregel** ist deterministisch (!)

$$\delta^*(\Omega_\kappa \mid \boldsymbol{x}) = \begin{cases} 1 & \text{falls } u_\kappa(\boldsymbol{x}) = \min_\lambda u_\lambda(\boldsymbol{x}) \\ 0 & \text{sonst} \end{cases}$$

und verwendet die **Prüfgrößen** (Trennfunktionen)

$$u_\lambda(\boldsymbol{x}) = \sum_{\kappa=1}^{K} r_{\kappa\lambda} \cdot P(\Omega_\kappa) \cdot P(\boldsymbol{x} \mid \Omega_\kappa) \quad \text{für } \lambda = 1, \ldots, K$$



Übernächste Woche:

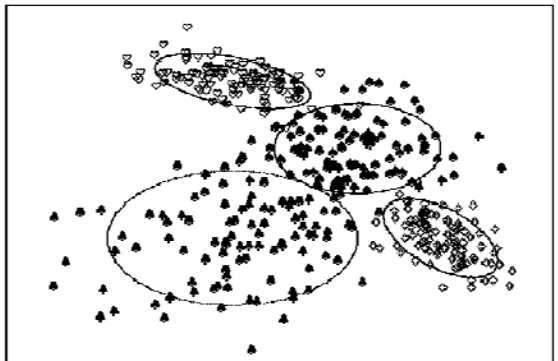
1.Overview Speech Recognition Systems &Architectures
2. Acoustic modeling & feature extraction (1)
3. Feature extraction (2)
4. LDA, Grundlagen der Klassifikation
5. Klassifikation in HMM Modellen
6. Wortmodellierung (trigramme, tying)
7. search/decoding, lattices,
   wordgraphs, confidence measures
8. acoustic adaptation
9. language models and grammars, Language model
   adaptation, lexica, phonology
10. speech understanding, dialogue control
11. Design of computer speech recognition systems

Statistische Clusteranalyse

ANNAHME: die Vektoren $\boldsymbol{x}$ sind **mischverteilt** gemäß

$$P(\boldsymbol{x} \mid \boldsymbol{p}, \boldsymbol{\theta}) \;=\; \sum_{\kappa=1}^{K} p_\kappa \cdot P(\boldsymbol{x} \mid \boldsymbol{\theta}_\kappa)$$

Bemerkung 4.2 *(Identifizierbarkeit von Mischverteilungen)*
Für gemischte *Normalverteilungen* sind die Parameterwerte $\boldsymbol{p}$, $\boldsymbol{\theta}$ eindeutig bestimmbar, sofern der genaue Funktionsverlauf von $P(\boldsymbol{x} \mid \boldsymbol{p}, \boldsymbol{\theta})$ bekannt ist!



Reelle Ebene mit 4 normalverteilten Clustern

Maximum-Likelihood-Schätzung

$$\mathcal{L}_{\text{MIX}}(\boldsymbol{p}, \boldsymbol{\theta}) \;=\; \log \prod_{i=1}^{N} P(\boldsymbol{y}_i \mid \boldsymbol{p}, \boldsymbol{\theta}) \;=\; \sum_{i=1}^{N} \log \left(\sum_{\kappa=1}^{K} p_\kappa P(\boldsymbol{y}_i \mid \boldsymbol{\theta}_\kappa) \right)$$

Gekoppelt-transzendente Gleichungen für $\hat{p}_\kappa$, $\hat{\boldsymbol{\mu}}_\kappa$, $\hat{\boldsymbol{\Sigma}}_\kappa$

$$\hat{p}_\kappa \;=\; \frac{1}{N} \sum_{i=1}^{N} P(\Omega_\kappa \mid \boldsymbol{y}_i, \hat{\boldsymbol{p}}, \hat{\boldsymbol{\theta}})$$

Estimation – Maximization (EM-Algorithmus)