4 Merkmalsextraktion

Wichtig : anhand der Merkmale eines Musters wird klassifiziert

- Merkmale müssen die *charakteristische* Information von Mustern enthalten (NICHT: gute Approximation oder Rekonstruktion)
- die Klassifikation soll “einfacher” sein
  - weniger Daten
  - kompakt im Merkmalsraum
- heuristische Merkmalsgewinnung
- analytische Merkmalsgewinnung
- nicht in der Vorlesung: Bewertung und Auswahl
- im wesentlichen numerische Merkmale: $\vec{f} \rightarrow \vec{c} \in R^n$

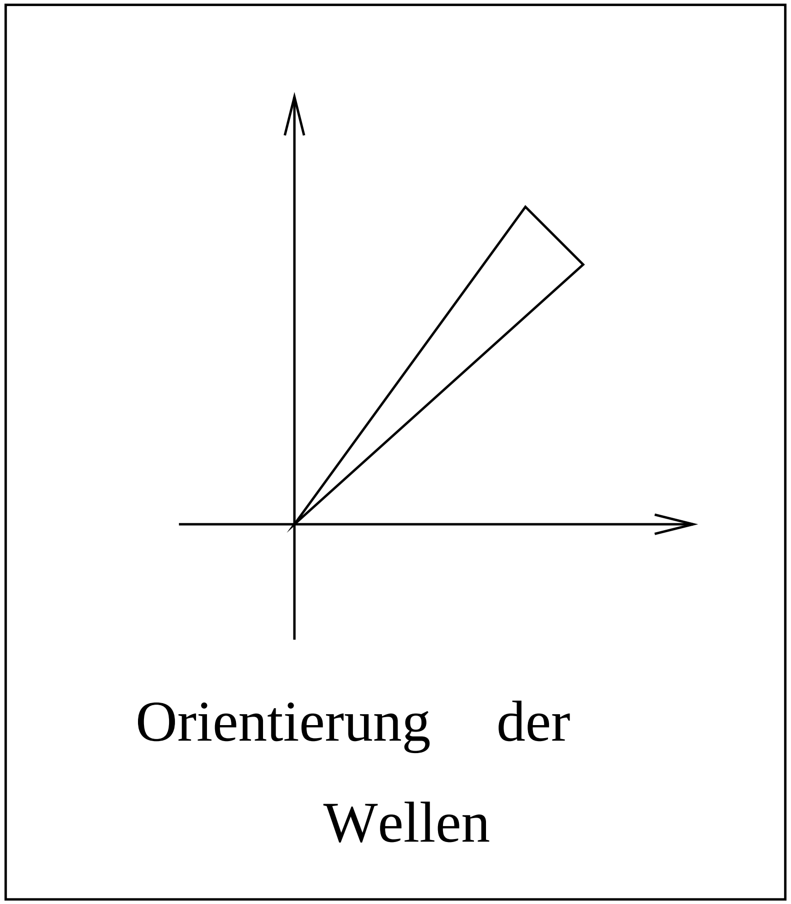
4.1 Heuristische Merkmale

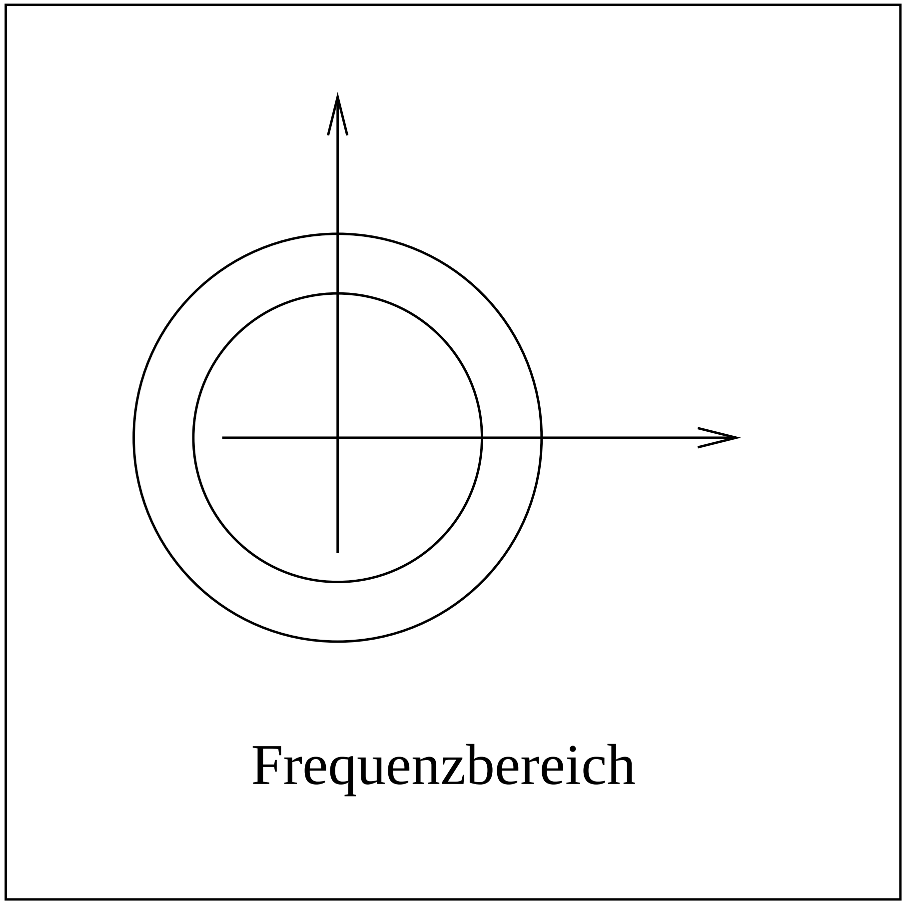
4.1.1 Kennzahlen

s. Abschnitt 2.1

4.1.2 DFT

- DFT ist translationsinvariant wenn nur die Beträge des Spektrums betrachtet werden
- Skalierung- oder Rotationsinvarianz durch geeignete vorherige Transformation
  z.B. Normierung mit Momenten (nur "Teil" davon)
- Reduktion der Anzahl von Merkmalen:
  Summiere die Koeffizienten der DFT auf gewissen Gebieten im Frequenzraum.
  Dabei enstehen eventuell Invarianzen
  Beispiele: Orientierung der Schwingungen, Frequenzbereich


Orientierung der
Wellen


Frequenzbereich

4.1.3 Lineare Vorhersage

Idee: schätze aus letzten m Abtastwerten den nächsten Abtastwert als Linearkombination:

$$\widehat{f_n} = -\sum_{\mu=1}^{m} a_\mu f_{n-\mu}$$

- damit kann der menschliche Vokaltrakt modeliert werden
- auch Anwendung z.B. für Textur

die a_μ werden so bestimmt, daß in einem Fenster von n_0 bis n_1 ein minimaler quadratischer Fehler entsteht, wir nehmen also an, daß das System in diesem Fenster annähernd stabil ist.

$$\epsilon = \sum_{n=n_0}^{n_1} (f_n - \widehat{f_n})^2$$

minimal wird.D.h.:

$$
\begin{aligned}
0 &= \frac{\partial \epsilon}{\partial a_\nu} \\
&= 2\left(\sum_{n=n_0}^{n_1}\left(f_n + \sum_{\mu=1}^{m} a_\mu f_{n-\mu}\right) f_{n-\nu}\right) \\
\sum_{\mu=1}^{m} a_\mu \sum_{n=n_0}^{n} f_{n-\mu} f_{n-\nu} &= -\sum_{n=n_0}^{n_1} f_n f_{n-\nu}
\end{aligned}
$$

Hinweis: Autokorrelationsmethode zum effizienten Bestimmen der a_ν

Für (oft überlappende) Fenster die a_μ bestimmen, daraus als Merkmale:

- die a_μ direkt, oder
- logarithmisches Modellspektrum als gegelättete Form des logaritmischen Spektrums des Musters $(\log |F_\nu|^2)$

$$\vec{a} = (1, a_1, \ldots, a_m, 0, \ldots, 0)$$

wird Fourier-Transformiert, dann logarithmieren, ergibt geglättetes Spektrum, eben das Modellspektrum.

Sprache typische Werte: $m = 10 \cdots 12$, $n_1 - n_0 \sim 300$

Die Länge M von $\vec{a}$ wird so gewählt, daß:

$$M > \frac{f_s}{f_r} = \frac{\text{Abtastfrequenz}}{\text{gewünschte Auflösung des Modellspektrums}}$$

4.1.4 Momente

Zentralmomente (mittelwertfrei)

- diskrete Version, Muster der Größe $M \times M$

$$\mu_{pq} = \sum_{k=0}^{M-1} \sum_{i=0}^{M-1} (x_k - x_s)^p \, (y_i - y_s)^q \, \Delta x \Delta y f_{ki}$$

$\Delta x, \Delta y$ Abtastintervalle

(x_s, y_s) Schwerpunkt

$\Rightarrow$ translationsinvarianz

Rotationsinvarianz

aus den Zentralmomenten, z.B.

$c_1 = \mu_{20} + \mu_{02}$

$c_2 = (\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2$

4.2 Analytische Merkmalsextraktion

- gewinne Merkmale c_ν *systematisch*, so daß (im gewissen Sinn) optimal
  Optimalität von “Problem”, d.h. den gegebenen Daten, abhängig
- wir betrachten lineare und orthonormale Transformationen:
  $\vec{c} = \Phi \vec{f}$, wobei i.d.R. $\dim \vec{c} < \dim \vec{f}$
  Zeilen von Φ orthonormal

Problem:

- bestimme Optimalitätskriterium
- bestimme optimale $\vec{\Phi}$

Zwei Typen von Kriterien:

1. Leistung des Klassifikators (z.B. Fehlerwahrscheinlichkeit) optimieren
   (gut, aber sehr aufwendig)
2. Postulat 3 (keine Rücksicht auf den Klassifikator, der spielt aufgrund des modularen Aufbaus keine Rolle)

Wir suchen n orthonormale Vektoren $\vec{\varphi}_\nu$ die eines der folgenden Kriterien erfüllen.

Wir gehen von einer Stichprobe aus

$$\omega = \left\{ {}^1\vec{f}, \ldots, {}^N\vec{f} \right\}$$

die Klassifiziert ist:

$$\begin{aligned} \omega &= \omega_1 \dot{\cup} \ldots \dot{\cup} \omega_k \\ \omega_\kappa &= \left\{ {}^1\vec{f}_\kappa, \ldots, {}^{N_\kappa}\vec{f}_\kappa \right\} \end{aligned}$$

Merkmale: ${}^i\vec{c}_\kappa$ ist der Merkmalsvektor zu ${}^i\vec{f}_\kappa$

Vier Kriterien (basierend auf euklidischem Abstand)

1. mittlerer quadratischer Abstand aller Merkmale untereinander

$$S_1 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left({}^i\vec{c} - {}^j\vec{c} \right)^T \left({}^i\vec{c} - {}^j\vec{c} \right)$$

2. Interklassen-Abstand

$$S_2 = \frac{1}{2 \cdot K \cdot (K-1)} \sum_{\kappa=2}^{K} \sum_{\lambda=1}^{\kappa-1} \frac{1}{N_\kappa N_\lambda} \sum_{i=1}^{N_\kappa} \sum_{j=1}^{N_\lambda} \left({}^i\vec{c}_\kappa - {}^j\vec{c}_\lambda\right)^T \left({}^i\vec{c}_\kappa - {}^j\vec{c}_\lambda\right)$$

Mittel über den mittleren quadratischen Abstand der Merkmale aus verschiedenen Klassen

3. Intraklassen-Abstand

$$S_3 = \frac{1}{K} \sum_{\kappa=1}^{K} \frac{1}{N_\kappa^2} \sum_{i=1}^{N_\kappa} \sum_{j=1}^{N_\kappa} \left({}^i\vec{c}_\kappa - {}^j\vec{c}_\kappa\right)^T \left({}^i\vec{c}_\kappa - {}^j\vec{c}_\kappa\right)$$

Mittlerer quadratischer Abstand von Merkmalen innerhalb einer Klasse

4. Kombination aus S_2, S_3 (S_2 groß, S_3 klein)

$$S_4 = S_2 - \lambda S_3$$

eine Möglichkeit:

- lege λ fest
- betrachte λ als Lagrange-Multiplikator

Suche $\vec{\Phi}$. so daß

- S_1, S_2 oder S_4 maximal werden oder
- S_3 minimal wird

Satz die $\vec{\Phi}^{(l)}$, die S_l optimieren, erhalten wir aus Eigenvektoren mit größten (bzw. kleinsten) Eigenwerten zu geeigneten *symmetrischen Kernmatrixen* $\vec{Q}^{(l)}$

$$\vec{Q}^{(1)} = \vec{R} - \vec{m}\vec{m}^T \quad \text{(Kovarianzmatrix)}$$

$$\vec{R} = \frac{1}{N} \sum_{j=1}^{N} {}^j\vec{f}\,{}^j\vec{f}^T \quad \text{(Korrelationsmatrix)}$$

$$\vec{m} = \frac{1}{N} \sum_{j=1}^{N} {}^j\vec{f} \quad \text{(Mittelwert)}$$

$$\vec{Q}^{(2)} = \frac{1}{k} \sum_{\kappa=1}^{k} \vec{R}_\kappa - \frac{2}{k(k-1)} \sum_{\kappa=2}^{k} \sum_{\lambda=1}^{\kappa-1} \left(\vec{m}_\kappa \vec{m}_\lambda^T + \vec{m}_\lambda \vec{m}_\kappa^T \right)$$

$$\vec{R}_\kappa = \frac{1}{N_\kappa} \sum_{j=1}^{N_\kappa} {}^j\vec{f}_\kappa\,{}^j\vec{f}_\kappa^T$$

$$\vec{m}_\kappa = \frac{1}{N_\kappa} \sum_{j=1}^{N_\kappa} {}^j\vec{f}_\kappa$$

$$\vec{Q}^{(3)} = \frac{1}{K} \sum_{\kappa=1}^{k} \left(\vec{R}_\kappa - \vec{m}_\kappa \vec{m}_\kappa^T \right)$$

$$\vec{Q}^{(4)} = \vec{Q}^{(2)} - \lambda \vec{Q}^{(3)}$$

Einschub

1. positiv definite Martrix $\vec{A}$ $\vec{x}^T\vec{A}\vec{x} > 0 \quad \forall\vec{x} \neq 0$
   für symmetrische, positiv definite Martrix $\vec{A}$ gilt:
   - $\vec{x}^T\vec{A}\vec{x}$ wird maximal für $\vec{x} =$ Eigenvektor zum größten Eigenwert von $\vec{A}$
   - am zweit größten, für den Eigenvektor zum nächst kleineren Eigenwert, ...
   - entsprechend bekommen wir die kleinsten Werte für die kleinsten Eigenwerte
2. $\vec{B}^T \cdot \vec{B}$ ist symmetrisch
3. für beliebige Martix $\vec{B}$ gilt

$$\vec{x}^T\vec{B}\vec{y} = \text{SP}\left(\vec{B}\vec{y}\vec{x}^T\right), \text{mit SP}\left(\vec{B}\right) = \text{Summe über die Diagonale}$$

4. für $M \times n$ Matrix $\vec{\Phi}$, $\vec{\Phi} = \left(\vec{\phi}_1, \ldots, \vec{\phi}_n\right)$, $M \times M$ Matrix $\vec{A}$ gilt

$$\text{SP}\left(\vec{\Phi}^T\vec{\Phi}\vec{A}\right) = \sum_{\nu=1}^{n} \vec{\phi}_\nu^T\vec{A}\vec{\phi}_\nu$$

damit **Beweis für** S_2

Zusammenfassung des Vorgehens

1. bestimme $Q^{(l)}$
2. bestimme die Eigenvektoren zu $Q^{(l)}$
   (Wenn $Q^{(l)}$ vollen Rang hat, besitzt sie M orthogonale Eigenvektoren)
3. wähle die n-Eigenvektoren zu den
   - größten (S_1, S_2, S_4)
   - kleinsten (S_3)

   Eigenwerten. Diese bilden die Matrix $\vec{\Phi}$

Bemerkungen:

1. diese Entwicklung heißt problemabhängig, da sie von der Stichprobe abhängt
   - die gewonnenen Merkmale enthalten in der Regel für die Klassifikation bessere Information als heuristische
   - sie sind meist näherungsweise normalverteilt
2. die Transformation, die S_1 maximiert heißt *Hauptachsentransformation* oder *Karhunen-Loeve-Transformation*

Eigenschaften der Hauptachsen Transformation:

- die c_ν sind maximal dekorreliert
- der mittlere quadratische Approximationsfehler wird minimal
  (bei fester Anzahl dim $\vec{c}$ Koeffizienten,
  wir können die Transformation in dim $\vec{c}$ dimensionale Unterraum auffassen!)
- die $\vec{\varphi}_\nu$ geben die Hauptachsen der Stichprobe, d.h. die Achsen ”größter Ausdehnung“
- häufige Modifikation

$$^j\widetilde{\vec{c}} = \vec{\Phi}\left({}^j\vec{f} - \vec{m}\right) = {}^j\vec{c} - \vec{\mu}$$