# Kapitel 2
# Prinzipien der Anfragebearbeitung in STMM-DBS

# 2. Prinzipien der Anfragebearbeitung in STMM-DBS

- Übersicht

# 2.1 Feature-Räume

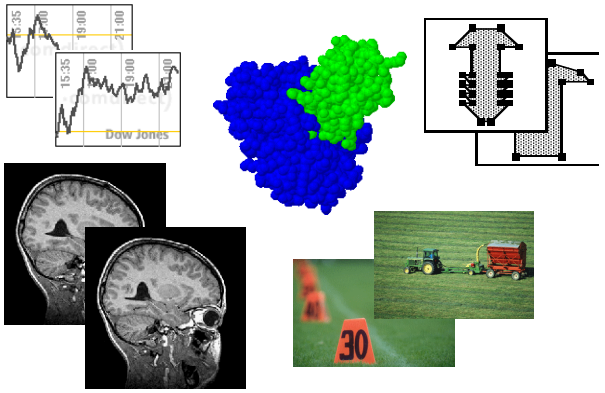
2.1 Feature-Räume

## *2.1.1 Das Prinzip der feature-basierten Ähnlichkeit*

- Grundidee der Feature-Transformation
  - Erwünscht: effiziente Ähnlichkeitssuche in Datenbanken
  - Meist ist Effizienz ohne Einsatz von Indexstrukturen nicht zu verwirklichen
  - Entwickle nicht für jedes der einzelnen Anwendungsgebiete/ Ähnlichkeitsmaße spezielle Indexstrukturen
  - Versuche mit wenigen Arten von Indexstrukturen möglichst viele Anwendungsgebiete abzudecken
  - Indexstrukturen für:
    - Multidimensionale Vektoren
    - Allgemein metrische Daten (beliebige Objekte, auf denen eine metrische Distanzfunktion definiert ist)

# 2.1.1 Das Prinzip der feature-basierten Ähnlichkeit

- Extrahiere charakteristische (nummerische) Eigenschaften („Features“) aus den Objekten



Anwendungsgebiete (Objektraum)

Feature-Transformation

Histogramm-Bildung

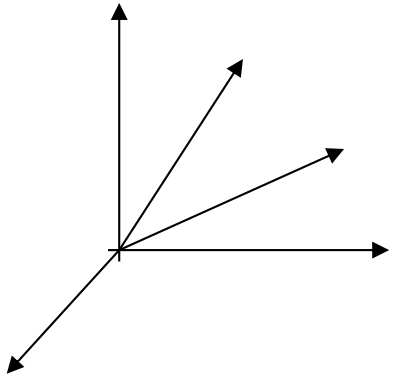
Moment-Invarianten

Rechteck-Abdeckung

Sektoren-Ermittlung

Fourier-Transformation Kurvatur

usw.



Featurevektor-Raum

- Wichtigste Eigenschaft der Feature-Transformation:
  - Ähnlichkeit der Objekte entspricht geringem Abstand der Feature-Vektoren

    => Ähnlichkeitsanfragen im Objektraum entsprechen Nachbarschaftsanfragen im Feature-Raum

  => Unterstützung durch geeignete multidimensionale Indexstrukturen

# 2.1.1 Das Prinzip der feature-basierten Ähnlichkeit

- Erweiterungen
  - Oft reichen die Mächtigkeit von Vektoren für die Modellierung nicht aus (vergleiche z.B. Relationales und OO-Modell)
  - Transformation in andere Räume, die die Definition einer Distanzfunktion mit Metrik-Eigenschaften erlauben
    - Graphen
    - Punktmengen
    - …
- Fazit:
  - Das Prinzip der Feature-Transformation ist ein mächtiges Werkzeug zur Modellierung der Ähnlichkeit von komplexen STMM-Objekten
  - Herausforderung: Finden geeigneter Feature-Transformationen

# 2.1.2 Feature-Räume und Distanzen

*2.1.2 Feature-Räume und Distanzen*

- Allgemeiner Feature-Raum

  Ein Feature-Raum ist ein Tupel $\Phi$ = (Dom, dist) mit

  - Dom ist ein Wertebereich (Domain)
  - dist ist eine Distanzfunktion, d.h. es gilt
    - Reflexivität $\forall x,y \in Dom: dist(x,y) = 0 \Leftrightarrow x=y$
    - Positiv-Definitheit $\forall x,y \in Dom, x \neq y: dist(x,y) > 0$
    - Symmetrie $\forall x,y \in Dom: dist(x,y) = dist(y,x)$

- Metrischer Raum

  Ein metrischer Raum ist ein Tupel $\Phi_M$ = (Dom, dist) mit

  - (Dom,dist) ist ein allgemeiner Feature-Raum
  - dist erfüllt zusätzlich die
    - Dreiecksungleichung $\forall x,y,z \in Dom: dist(x,z) \leq dist(x,y) + dist(y,z)$

# 2.1.2 Feature-Räume und Distanzen

- Euklidischer Vektorraum

  Ein (euklidischer) Vektorraum der Dimension *d* (*d*-dimensionaler Vektorraum) ist ein Tupel $\Phi_E$ = (Dom, dist) mit

  - (Dom,dist) ist ein metrischer Raum
  - Dom = $\mathbb{R}^d$

- Feature-Transformation

  Eine Feature-Transformation ist eine Abbildung

  $$T: \text{OBJ} \rightarrow (\text{Dom},\text{dist})$$

  die jedem Objekt $o \in \text{OBJ}$ aus dem Objektraum ein Objekt aus dem Wertebereich Dom zuordnet.

  Die Distanz im Objektraum wird durch die Distanz im Feature-Raum repräsentiert, d.h.

  $\forall x,y \in \text{OBJ}: \text{dist}_{\text{OBJ}}(x,y) \equiv \text{dist}_{\text{Dom}}(T(x),T(y))$

# 2.1.2 Feature-Räume und Distanzen

- Distanzmaße in Vektorräumen
  - **Euklidische Norm ($L_2$):**

    $dist = ((p_1-q_1)^2+(p_2-q_2)^2+...)^{1/2}$

    Natürliches Distanzmaß

  - **Manhattan-Norm ($L_1$):**

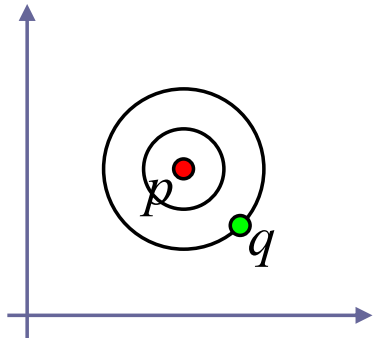
    $dist = |p_1-q_1|+|p_2-q_2|+...$

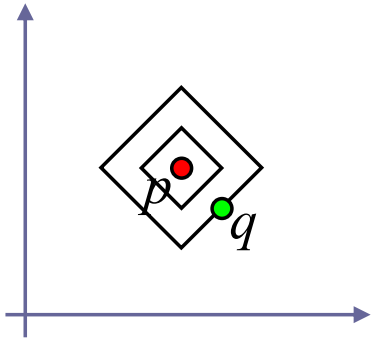
    Die Unähnlichkeiten der einzelnen Merkmale werden direkt addiert

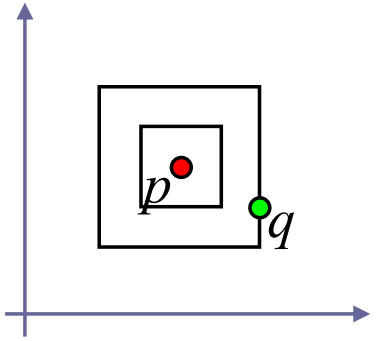
  - **Maximums-Norm ($L_\infty$):**

    $dist = \max\{|p_1-q_1|, |p_2-q_2|,...\}$

    Die Unähnlichkeit des am wenigsten ähnlichen Merkmals zählt



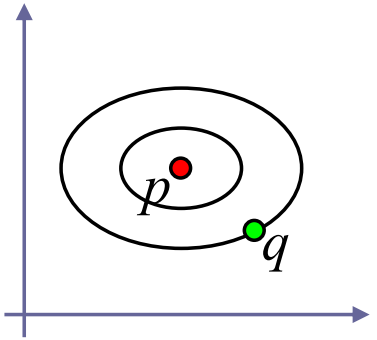



# 2.1.2 Feature-Räume und Distanzen

- Verallgemeinerung $L_p$-Abstand: $dist_p = (|p_1-q_1|^p + |p_2-q_2|^p + ...)^{1/p}$
- **Gewichtete Euklidische Norm:**

  $dist = (w_1(p_1-q_1)^2+w_2(p_2-q_2)^2+...)^{1/2}$

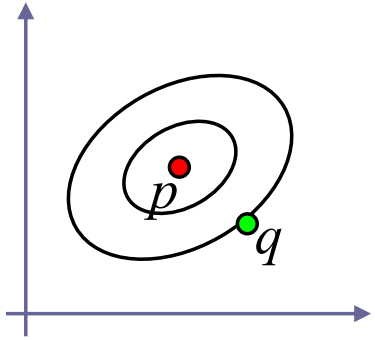
  Häufig sind die Wertebereiche der Merkmale deutliche unterschiedlich
  Beispiel: Merkmal $M_1 \in [0.01 \ldots 0.05]$
  Merkmal $M_2 \in [3.07 \ldots 22.2]$
  Damit M1 überhaupt berücksichtigt wird muss es höher gewichtet werden




- **Quadratische Form:**

  $dist = ((p - q)\ \mathbf{M}\ (p - q)^T)^{1/2}$

  Bisherige Abstandsmasse gewichten Merkmale nur getrennt
  Besonders bei Farbhistogrammen müssen verschiedene Merkmale gemeinsam gewichtet werden



# 2.1.2 Feature-Räume und Distanzen

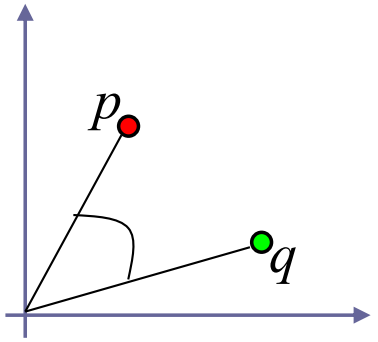
- **Cosinus-Distanz**

  *dist* = cos(winkel(p,q))

  Berechnet den cosinus des Winkels

  Meist für sehr hochdimensionalen

  Featurevektoren (z.B. Texten)




- Bemerkungen
  - Jeder Vektorraum ist ein metrischer Raum, jeder metrische Raum ein allgemeiner Feature-Raum
  - Sprechweise meist: „Feature-Raum“ statt (euklidischer) Vektorraum
  - Transformation komplexer Objekte meist immer in metrische Räume wegen der Dreiecksungleichung (Performanz!!!)