Fallbasiertes Schließen, FBS bzw. CBR = Case Based Reasoning

1. Grundidee: Handeln auf der Basis von Erfahrungen

> Jede von mir gelöste Aufgabe wurde zum Muster, welches ich im weiteren für die Lösung anderer Aufgaben benutzte.
>
> DESCARTES

> Der Mensch hat dreierlei Wege klug zu handeln:
>
> Erstens durch **Nachdenken**,
> – das ist der edelste,
>
> zweitens durch **Nachahmen**,
> – das ist der leichteste,
>
> und drittens durch **Erfahrung**,
> – das ist der bitterste.
>
> KONFUZIUS

> Alles Wissen stammt aus Erfahrung.
>
> KANT

Problem:
Das Bild des Beamers ist unscharf und flackert

Lösungsversuche:
... adjustieren gemäß Angaben der Bedienungsanleitung
... Zeilenfrequenzen ändern
... Auflösung ändern
... Kabel untersuchen

Problem:
Das Bild des Beamers ist unscharf und flackert

Lösungsversuche:
... adjustieren gemäß Angaben der Bedienungsanleitung
... Zeilenfrequenzen ändern
... Auflösung ändern
... Kabel untersuchen
... **Stecker umdrehen!**

Fall = Problem + Lösung: $\boxed{\text{P}} + \boxed{\text{L}}$

$\boxed{\text{P}} = \boxed{\text{Das Bild des Beamers ist unscharf und flackert.}}$
$\boxed{\text{L}} = \boxed{\text{Stecker umdrehen.}}$

Grundschema (naive Sicht):

Neues Problem P: unbekannte Lösung L.
Altes Problem P': bekannte Lösung L'.
$\Rightarrow$ Alte Lösung L' an Problem P anpassen

$$\begin{tikzcd}
\boxed{P} \arrow[rr] \arrow[d, dash, "\sim"] & & \boxed{L\ ?} \arrow[d, dash, "\sim"] \\
\boxed{P'} \arrow[rr] & & \boxed{L'}
\end{tikzcd}$$

Ausgangs-Hypothese (später genauer untersuchen):

„Ähnliche Fälle haben ähnliche Lösungen“

A-posteriori-Ähnlichkeit:

$$\boxed{L_1} \sim \boxed{L_2} \quad \Rightarrow \quad \boxed{P_1} \sim \boxed{P_2}$$

i.a. nicht vernünftig realisierbar
(Bsp.: gleiche Strafe für unähnliche Taten).

Umkehrung:

$$\boxed{L_1} \sim \boxed{L_2} \quad \Leftarrow \quad \boxed{P_1} \sim \boxed{P_2}$$

Notwendig bei normativen Problemen (Rechtsprechung).
Falls nicht erfüllt: Ähnlichkeit falsch definiert?

Extremfall: Ähnlichkeit = Identität

Später: graduelle Ähnlichkeit (Ähnlichkeitsmaße)

Kognitiver Aspekt der KI

Modellieren der natürlichen Intelligenz, Imitation, Experimente

Technischer Aspekt der KI

Automatisierung geistiger Prozesse, „Intelligente“ Maschinen, Programme, ...

KI-Ansätze:

- Logik
- Stukturierte Repräsentationen
- Regeln
- Constraint Propagation
- Suchverfahren
- Neuronale Netze
- ...
- Fallbasiertes Schließen

Lösungen „from scratch“ ?
Kognitive Adäquatheit.

Experten-Wissen: $\boxed{\textbf{Studium + Erfahrung}}$

> Human experts are not systems of rules,
> they are libraries of experiences.
>
> RIESBECK/SCHANK

„Allgemeines“ Wissen : Lehrbücher, Regeln, Gesetze, Normen, ...

„Spezielles“ Wissen : Erfahrungen: „Fälle“

Ich hatte da einen Fall ...
In solchen Fällen muß man ...

- Medizin
- Rechtssprechung
- Technische Diagnose
- Konstruktion (⇒ „Motorkutsche“)
- ...
- Kreativität, Intuition, ...

Denken in Fällen

zur Erklärung : Fallbeispiel
ein Beispiel sagt mehr als tausend Erklärungen

zur Aufstellung von Regeln – Generalisierung :
Experimente (Fälle) zur Erzeugung, Überprüfung
Regelfall, allgemeiner, abstrakter, typischer Fall, ...

zur Eingrenzung von Regeln :
Ausnahmefall, Sonderfall, Einzelfall, ...

zur Übertragung – Analogien :
analoger Fall, Vorbild, ...

zum Lernen : Probefall, Fehlerfall, ...
aus Fehlern lernen

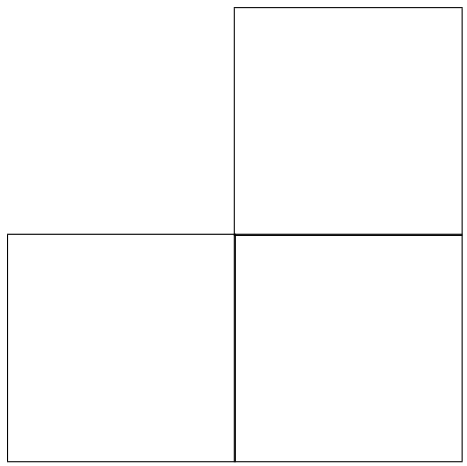
Unterschiedliche Abstraktionsebenen:
„typischer Fall“ als konkretes Beispiel („Element“) für einen „Regelfall“ („Menge“).

Generalisierung, Abstraktion:
Übergang zu allgemeinen Aussagen (z.B. Regeln)
daneben stets: *Ausnahmefall, Sonderfall, Einzelfall, ...*

Fälle in der Umgangssprache:

Ausnahmefall,
Spezialfall,
Parallelfall,
Einzelfall,
Zwischenfall,
Vorfall,
Bedarfsfall,
Wiederholungsfall,
Präzedenzfall,
Grenzfall,
Glücksfall,
Zweifelsfall,
Ausfall,
Zufall,
Wasserfall,
Unfall,
Überfall,
Anfall,
Einfall,
Sündenfall,
Abfall

Aufgabe: Figur in 4 kongruente Teile zerlegen



Container: Grundbestandteile eines Fallbasierten Systems:

- Vokabular (Falldatenbasis)
- Ähnlichkeit von Fällen
- Auswahlmechanismus (Retrieval)
- Transformation: Anpassung von Lösungen

„**Ähnlich**“ oder „analog“:

Fallbasiertes Schließen: innerhalb einer Domäne
Analoges Schließen: Übertragung auf andere Domäne

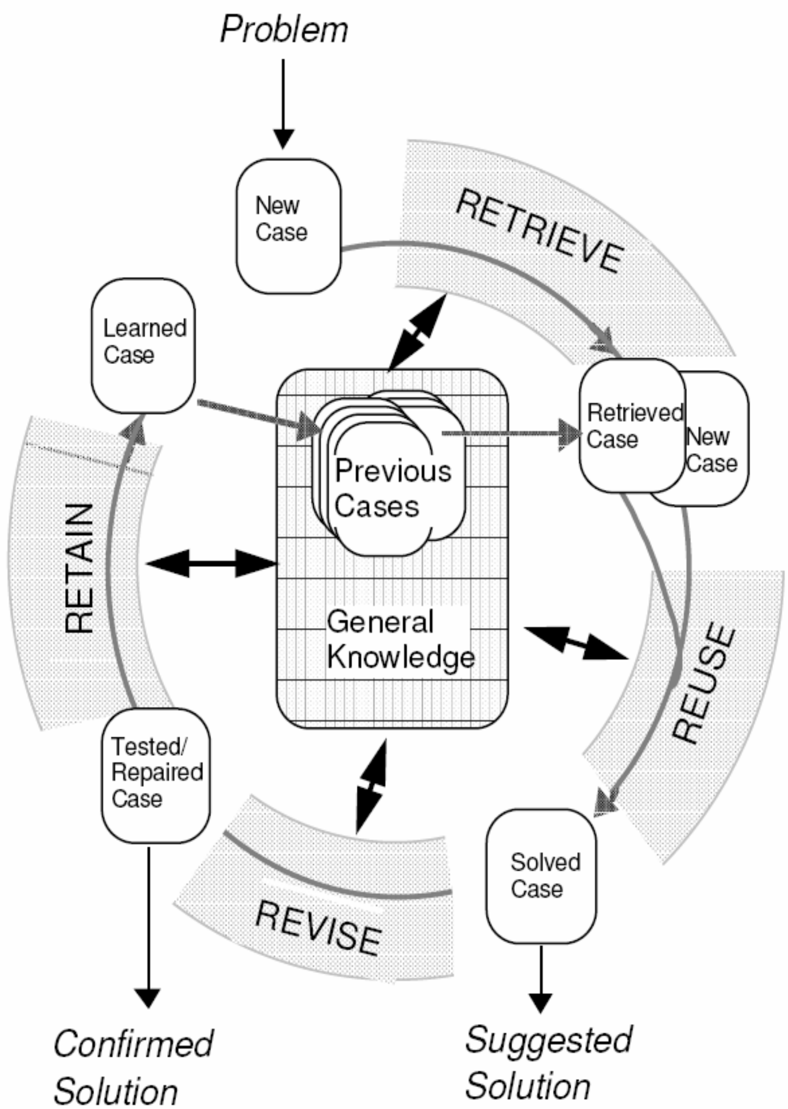
Klassische **Falldarstellung**: Problem + Lösung (+ Erklärung) (+ Bewertung)

Attribut-Werte-Paare, Texte, Grafiken, Bilder, ...
Datenbanken, Protokolle, Log-Files, Notizen, ...

Fallformate:

Stark strukturiert :
z.B. Immobilien, Produkt-Kataloge, ... (Datenbanken)

Schwach bzw. heterogen strukturiert :
z.B. Fehlerprotokolle, Patientenakten, Baupläne, ...



R4-Modell von Aamondt und Plaza

Zyklus:

1. **Aufbereitung des aktuellen Problems**
   Beschreibungsform anpassen
   Bestimmung von Merkmalen, Indizes
2. **Suche nach „relevanten/ähnlichen Fälle“**
   mittels Indizierung, Ähnlichkeit, ...
3. **Auswahl der „brauchbaren Fälle“**
   im Sinne „anwendbarer Erfahrungen“ (auch negativ)
4. **Konstruktion einer Lösung (Anpassung)**
   „Argumentieren“ mit den vorliegenden Fällen,
   Anpassung früherer Lösungen an das neue Problem
5. **Kritik der Lösung(en)**
   Erneuter Vergleich mit der Falldatenbasis,
   Simulation, ggf. weitere Anpassungsschritte
6. **Praktische Anwendung**
   Analyse und Bewertung
7. **Aktualisierung der Falldatenbasis**
   Einordnen des neuen Falls (Wissensakquisition)

Pflege der Falldatenbasis:

- Streichen („Vergessen“) irrelevanter Fälle
- Reduktion: Redundante Fälle streichen
- Korrektur der Auswahlkriterien (Ähnlichkeit)

2. Klassifikationsprobleme

- Menge O von Objekten $o_1, o_2, ...$
- Menge $\mathcal{K}$ von Klassen $K_1, K_2, ...$ (jeweils $K_i \subseteq O$), Zusatzforderung: $K_i \cap K_j = \emptyset$ für $i \neq j$

$$K(o) := \text{ (korrekte!) Klasse von } o.$$

- Klassifikator:

$$\kappa : O \to \mathcal{K}$$

- korrekter Klassifikator:

$$\forall o \in O : \ \kappa(o) = K(o) \qquad \text{d.h.: } o \in \kappa(o)$$

Merkmale (Attribute, Features, ...)

- Mehrere Merkmale $m_1, m_2, ..., m_n$ (Merkmalsmenge) jeweils über einer Wertemenge W_i
- Merkmale m_i ordnen den Objekten Werte aus einer Wertemenge W_i zu:

$$m_i : O \to W_i \qquad \text{(evtl. partielle Funktion)}$$

- Merkmalsraum: $M := W_1 \times ... \times W_n$
- jedem Objekt $o \in O$ ist Merkmalsvektor zugeordnet (evtl. einige $m_i(o)$ nicht definiert):

$$m(o) := [m_1(o), ..., m_n(o)]$$

Anforderung an M: $m(o') = m(o'') \Rightarrow o' = o''$
oder etwas schwächer: $m(o') = m(o'') \Rightarrow \kappa(o') = \kappa(o'')$

Klassifizierungsaufgabe neu formulieren:
gesucht wird eine Klassifikator $\kappa : M \to \mathcal{K}$ mit der Korrektheitsforderung $\forall o \in O : \ o \in \kappa(m(o))$

Beobachtungen:

1. Unterschiedliche Relevanz von Merkmalen

   Extreme Varianten:

   irrelevantes Merkmal: $m_i(o') = m_i(o'')$ für alle $o' \neq o''$

   global unterscheidendes Merkmal $m_i(o') \neq m_i(o'')$ für alle $o' \neq o''$

   i.a. nur Kombinationen von Merkmalen ausreichend für korrekte Klassifizierung

2. Redundanze: $\Rightarrow$ Auswahl von Merkmalen

3. Clusterbildung:
   Cluster = Objekte mit "**ähnlichen**" Merkmalen (z.B. "benachbart" bzgl. einer Distanz)

   abhängig von Merkmalsauswahl/Ordnungen über W_{m_i}
   Ziel: **Ähnlichkeit** (Distanz) entspricht Klassen: $m(o') \sim m(o'') \Rightarrow K(o') = K(o'')$

4. Evtl. bestehen Klassen aus mehreren Clustern.

$\Rightarrow$ Weitere Anforderungen an Merkmalsauswahl.

Klassifizierungsverfahren

Komplexitätsprobleme:
a) bzgl. Konstruktion des Klassifizierungsverfahrens κ
b) bzgl. Ausführen einer Klassifikation $\kappa(o)$
Evtl. $\kappa(o)$ nur näherungsweise korrekt.

Im folgenden zunächst: $\boxed{\text{Ähnlichkeit als inverse Distanz}}$

Trennflächen : *Trennung der Klassen im Merkmalsraum M durch Trennflächen*

lineare Trennbarkeit: Klassen durch lineare Trennflächen separiert

Entscheiden, in welchem Teil des Raumes ein Merkmalsvektor $[m_1(o), ..., m_n(o)]$ liegt.

Beispiel:

Trennlinie $a \cdot x + b \cdot y + c = 0$

Klassifizierung für(x_0, y_0) gemäß $a \cdot x_0 + b \cdot y_0 + c \geq 0$ bzw. $a \cdot x_0 + b \cdot y_0 + c < 0$

Support-Vektor-Maschinen

- Günstige Lage der Trennebene (gute Trennbarkeit)
- Transformation des Raumes für lineare Trennbarkeit

Nearest Neighbour Verfahren: Nähe zu bekannten Objekten

Gegeben eine Menge C von bekannten Objekten ("Fällen"):

$$C = \{\ c_j = [m(c_j), \kappa(c_j)] \mid j = 1, \ldots, r\}$$

Nearest Neighbour Verfahren:

$$\kappa(o) := \kappa(c_j) \text{ für } c_j := NN(o, C)$$

wobei

$$\begin{aligned} NN(o, C) &:= \textit{ähnlichstes Objekt } c \in C \textit{ für } o \\ (bzw. &:= \textit{Objekt } c \in C \textit{ mit minimaler Distanz zu } o) \end{aligned}$$

(Problem: Eindeutigkeit von NN)

Korrektheit abhängig von

– Korrektheit der Fälle in C
– Ähnlichkeitsmaß (Distanzmaß)
– Vollständigkeit, Verteilung der Fälle in C

(vgl. Diskussion „Ähnliche Fälle haben ähnliche Lösungen“)

Voronoi-Diagramm: Trennflächen gemäß Nearest-Neighbour-Klassifikation

n Nearest Neighbour Verfahren:

Klassifizierung entsprechend der Mehrzahl der n nächstliegenden Objekte.

Mittel-/Schwerpunkte von Klassen (Clustern)

Klassifizierung entsprechend dem nächstliegenden
Mittel-/Schwerpunkt:

Mittelpunkt einer Klasse

$K = \{[m_1(o_1), ..., m_n(o_1)], \ldots, [m_1(o_{|K|}), ..., m_n(o_{|K|})]\}$:

$$m_K := \frac{1}{|K|} \sum_{k=1,...,|K|} [m_1(o_k), ..., m_n(o_k)]$$

Verfahren:
Nearest Neighbor Verfahren mit Fallbasis C = Mittelpunkte

Analog für Schwerpunkte:
Schwerpunkt einer Klasse:
wie Mittelpunkt aber mit gewichteten Merkmalsvektoren

$$s_K := \frac{1}{|K|} \sum_{k=1,...,|K|} g_i \cdot [m_1(o_k), ..., m_n(o_k)]$$

(g_i = Gewicht von m_i)

Entscheidungsbäume als Klassifikatoren

Zu klassifizieren: o
Eingabe: $m(o) = [m_1(o), ..., m_n(o)]$

- Knoten mit Nachfolgern: Tests mit Verzweigung entsprechend Resultat.
  z.B. Test: IF $m_1(o) > a$ THEN rechter Zweig ELSE linker Zweig
- Blattknoten: Ausgabe $\kappa(o)$

Test/Verzweigung als *Abschneiden* des Merkmalsraums:
Suche nur noch im ausgewählten Unterraum.
Bei einfachen Tests (z.B. $m_i > a$) : Teilräume begrenzt durch achsenparallele Ebenen.

Beobachtungen:
Wenig Tests bei "guter" Clusterung.
Wenig Tests bei guter Anordung/Auswahl der Tests.

Klassifizierungsregeln
Erzeugbar z.B. aus Entscheidungsbäumen
(jeweils entlang der Pfade):
$\texttt{IF} \;\; m_i(o) > a_i \; \texttt{AND} \;\; m_j(o) > a_j \ldots \texttt{THEN} \;\; \kappa(o) = K$

Neuronale Netze als Klassifikatoren

Statistische Verfahren

Lernverfahren zur Erzeugung von Klassifikatoren (Maschinelles Lernen)

Ausgangspunkt: Es sind nur einige Klassifizierungsbeispiele bekannt.

positive Beispiele: $[[m_1(o), ..., m_n(o)], K]$ mit $o \in K$

evtl. auch
negative Beispiele: $[[m_1(o), ..., m_n(o)], K]$ mit $o \notin K$

Ziel: Klassifikator κ ableiten, der (möglichst) korrekt ist.

- Entscheidungsbäume: *ID3 und Nachfolger, speziell C4.5*
- Entscheidungsregeln aus Entscheidungsbäumen
- Neuronale Netze z.B. mittels Backpropagation

Generalisierungen beim Lernen

Beispielmenge C gegeben. Konstruiert wird Klassifikator κ *für alle* Objekte.

Probleme:

- Wonach bestimmt sich Klassifizierung für $o \notin C$?
- muss Klassifikator für Beispiele aus C korrekt sein?

Induktiver Bias: Für Verallgemeinerung benutzte Annahme.

Beispiel: Nearest Neighbor mit Fallbasis C

$\forall c \in C : c \in \kappa(c)$.
Für andere Objekte $o \notin C$: $\kappa(o) := \kappa(c) \qquad \text{für} \quad c := NN(o, C)$

$\Rightarrow$ *Generalisierung* gemäß Nähe. (*induktiver bias*)

Beispiel: Lineare Trennflächen
$\Rightarrow$ *Generalisierung* gemäß Trennflächen. (*induktiver bias*)

dabei evtl. Verzicht auf korrekte Klassifizierung aller Beispiele zwecks Vereinfachung der Teilräume (insbesondere könnten einige Beispiele fehlerhaft gewesen sein.)

3. Ähnlichkeit

„ähnliche Fälle haben ähnliche Lösungen“

> An essay is like a fish.
>
> TVERSKY

> Any two things which are from one point of view similar may be dissimilar from another point of view.
>
> POPPER

Eigentlich wichtig sind a-posteriori-Kriterien:

- Korrektheit der Klassifizierung
- Übertragbarkeit der Probleme/Lösungen
- Nützlichkeit der gefundenen Lösung
- Akzeptanz des Nutzers

Probleme:
– hoher Aufwand für Abschätzung
– Überprüfung erst nach Abschluss

Ersatz durch leichter faßbare a-priori-Kriterien:

- bzgl. Berechnung
- bzgl. Retrieval
- bzgl Transparenz

$\Rightarrow$ Ähnlichkeit

Betrachten im folgenden zunächst **Merkmalsvektoren** als Objekte:

$$\boxed{u \in U := \mathcal{R} \times \mathcal{R} \times ... \times \mathcal{R}}$$

Basis-Konzepte für Ähnlichkeit:

- Binäre Relation R_{SIM} über U
  ($R_{SIM}(u, v)$: "u ist ähnlich zu v"):

  $$R_{SIM} \subseteq U \times U.$$

  Beispiel:

  Übereinstimmung bei mindestens 2 von 3 Merkmalen:

$$\begin{aligned} &R_{SIM}([x_1, x_2, x_3], [y_1, y_2, y_3]) \\ &\quad :\Leftrightarrow \exists i, j : 1 \leq i < j \leq 3 : x_i = y_i \wedge x_j = y_j. \end{aligned} \tag{1}$$

- Ähnlichkeitsmaß SIM
  ($SIM(u,v)$: Grad der Ähnlichkeit zwischen u und v

$$SIM : U \times U \rightarrow S.$$

  mit $S \subseteq \mathcal{R}$, z.B. $S = [0,1]$.

  Beispiele

  – „*Simple matching coefficient*“:

$$SIM([x_1, x_2, x_3], [y_1, y_2, y_3]) := \frac{1}{3} \cdot card(\{\ i \mid x_i = y_i\}). \tag{2}$$

  – Maß abhängig von Differenzen $x_i - y_i$ z.B.:

$$SIM([x_1, x_2, x_3], [y_1, y_2, y_3]) := \frac{1}{3} \cdot \sum_{i=1,2,3} \frac{1}{1 + |x_i - y_i|}. \tag{3}$$

- Nachbarschaft bzgl Distanz $DIST$
  ($DIST(u,v)$: Distanz zwischen u und v):

$$DIST : U \times U \to S.$$

  Beispiel:
  Manhattan-Distanz über $S = [0, \infty)$:

$$DIST([x_1, x_2, x_3], [y_1, y_2, y_3]) := \sum_{i=1,2,3} |x_i - y_i|. \tag{4}$$

Wertebereiche für $S \subseteq \mathcal{R}$ bei $DIST$ bzw. SIM

$S = [0, 1]$ analog zu Stochastik, Fuzzy-Theorie etc.

- bei SIM : Maximaler Wert 1 : absolute Ähnlichkeit,
- bei SIM : Minimaler Wert 0 : absolute Unähnlichkeit,
- bei $DIST$: Minimaler Wert 0 : absolute Nähe,

$S = [0, \infty]$

- bei $DIST$ üblich
- bei SIM :Problematisch für Selbst-Ähnlichkeit: Kein maximaler Wert für $SIM(x, x)$

S mit negativen Werten:

- Negative Ähnlichkeiten $\Rightarrow$: Ablehnung
- negative Distanzen?

Auswirkungen bzgl. Retrieval-Strategien (s.u.)

Undefinierte/unbekannte Werte

2 Interpretationen:

- unwichtig, egal
  (z.B. fehlendes Attribut in einer Regel)
- unbekannt

Unterschiedliche Behandlung in Anfrage bzw. Fall bei FBS.

Möglichkeiten:

- Normierung über bekannte Werte, z.B.

$$dist(x,y) := \sqrt{\frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|^2}$$

- Optimistisch:
  Annahme: unbekannte Werte stimmen überein

$$\begin{aligned} &SIM([u_1,..u_{i-1},?,u_{i+1}.,u_n],[v_1,...,v_n]) \\ &\quad := Max_{u_i\in\mathcal{R}}\, SIM([u_1,..u_{i-1},u_i,u_{i+1}.,u_n],[v_1,...,v_n]) \end{aligned}$$

- Pessimistisch:
  Annahme: unbekannte Werte stimmen nicht überein

$$\begin{aligned} &SIM([u_1,..u_{i-1},?,u_{i+1}.,u_n],[v_1,...,v_n]) \\ &\quad := Min_{u_i\in\mathcal{R}}\, SIM([u_1,..u_{i-1},u_i,u_{i+1}.,u_n],[v_1,...,v_n]) \end{aligned}$$

Beziehungen

- R_{SIM} betrachten als SIM mit $S = \{0, 1\}$.

- R_{SIM} definierbar aus SIM mittels *Nearest Neighbor* unter bezug auf eine Fallmenge $C \subseteq U$:
  $R_{NN_C}(u, c) :\Leftrightarrow c = NN(u, C)$
  d.h.
  $$R_{NN_C}(u, c) :\Leftrightarrow \forall c' \in C : SIM(u, c) \geq SIM(u, c').$$
  Analog für Distanzen:
  $$NN_C(u, c) :\Leftrightarrow \forall c' \in C : DIST(u, c) \leq DIST(u, c').$$
  Bemerkung: Definitionsbereich hierbei $U \times C$.

- R_{SIM} definierbar aus SIM mittels „Mindest-Ähnlichkeit“ b:
  $$R_{SIM_b}(u, v) := SIM(u, v) > b.$$
  (Analog für Distanz mit $DIST(u, v) < b$.)

Vergleichsrelationen

Definitionen:

- *y ist ähnlicher zu x als v zu u* :
  4-stellige Relation $V4_{SIM}(x, y,\ u, v)$ mit

  1. $\forall x, u, v : V4_{SIM}(x, x,\ u, v)$
     (Reflexivität)
  2. $\forall x, y, u, v : V4_{SIM}(x, y,\ u, v) \Leftrightarrow V4_{SIM}(y, x,\ u, v)$
     $\Leftrightarrow V4_{SIM}(x, y,\ v, u)$
     (Symmetrie, evtl. verzichtbar, s.u.)

- *y ist ähnlicher zu x als z* :
  3-stellige Relation $V3_{SIM}(x, y, z)$ definierbar durch

  $$V3_{SIM}(x, y, z) := V4_{SIM}(x, y,\ x, z)$$

- **Präferenzrelation bzgl. fixiertem** x:
  2-stellige Relation $PREF_x(y, z)$ definiert durch

  $$PREF_x(y, z)\ :=\ V3_{SIM}(x, y, z)$$

  $PREF_x(y, z)$ ist Halbordnung,
  Maximale Elemente = Nearest Neighbor(s).
  Im FBS: x als Anfrage.

Kompatibilität

Definition

Gegeben SIM bzw. $DIST$

$$V4_{SIM}(x,y,u,v) \;:\Leftrightarrow\; SIM(x,y) \geq SIM(u,v),$$
$$V4_{DIST}(x,y,u,v) \;:\Leftrightarrow\; DIST(x,y) \leq DIST(u,v).$$

Zwei (Ähnlichkeits- oder Distanz-)Maße m_1 und m_2 sind **relational kompatibel** gdw.

$$\forall x,y,u,v \in U : V4_{m_1}(x,y,u,v) \Leftrightarrow V4_{m_2}(x,y,u,v).$$

(Evtl. nur auf Teilmengen von $U \times U \times U \times U$.)

Kompatible Transformation zwischen Maßen m_1 und m_2 durch Funktion f mit

1. Respektierung der Wertebereiche S_i der Maße m_i,
2. Respektierung der "Reflexivitätsbedingungen" (s.u.), z.B. $SIM(x,x) = 1$, $DIST(x,x) = 0$
3. Respektierung der Kompatibilitätsbedingung, z.B. durch eineindeutige Funktion

Beispiele:

Funktion f für Umwandlung

$$DIST : U \times U \rightarrow [0, \infty] \quad \Rightarrow \quad SIM : U \times U \rightarrow [0, 1]$$

f monoton fallend, $\quad f(0) = 1, \quad lim_{x \rightarrow \infty} f(x) = 0.$

z.B.:

$$f(x) = \frac{1}{1 + x}$$

Funktion f für Umwandlung

$$SIM : U \times U \rightarrow [0, 1] \quad \Rightarrow \quad DIST : U \times U \rightarrow [0, \infty]$$

f monoton fallend, $\quad f(1) = 0, \quad lim_{x \rightarrow 0} f(x) = \infty.$

z.B.:

$$f(x) := \frac{1 - x}{x}$$

Mögliche Annahmen (Axiome)
(z.B. für $S = [0, 1]$ bzw. $S = [0, \infty]$)

Reflexivität

$$SIM(x, x) = 1 \quad \text{und für Distanzen:} \quad DIST(x, x) = 0.$$

Symmetrie

$$SIM(x, y) = SIM(y, x) \quad \text{und} \quad DIST(x, y) = DIST(y, x).$$

Dreiecksungleichung

$$DIST(x, z) \leq DIST(x, y) + DIST(y, z).$$

(Bzgl. Ähnlichkeit: s.u.)

Einzige ständige Voraussetzung im folgenden:
Reflexivität in der Form $SIM(x, x) = 1$ bzw. $DIST(x, x) = 0$.

Reflexivität

$SIM(x, x) = 1$ bzw. $DIST(x, x) = 0$. sichert: $c = NN(c, C)$ für $c \in C$

Evtl. schärfer: „*starke Reflexivität*“

$$SIM(x, y) = 1 \leftrightarrow x = y \quad \text{und} \quad DIST(x, y) = 0 \leftrightarrow x = y.$$

(z.B. bei Metriken)

Symmetrie

Im FBS: Vergleich von Anfragen $q \in U$ mit Fällen $c \in C$ (bzw. mit „Problembeschreibungen“).
Bsp.: Nearest neighbor.
$\Rightarrow$ unsymmetrische Definitionsbereiche: SIM bzw. $DIST$ definiert auf $U \times C$

Unsymmetrie in Alltagssprache („gerichtete Vergleiche“):

- Tochter ähnlich zur Mutter,
- Pinguin ähnlich zu Vogel,
- Anfrage ähnlich zu Dokument

Distanzen (speziell Metriken): symmetrisch.

Transitivität, Dreiecksungleichung

Ähnlichkeits-Relationen R_{SIM} i.a. nicht transitiv.
In Beispiel (1):
$R_{SIM}([1,1,1],[1,1,0])$ und $R_{SIM}[1,1,0],[1,0,0])$, aber nicht $R_{SIM}([1,1,1],[1,0,0])$.

Dreiecksungleichung für Ähnlichkeitsmaße?

Für Distanzen gilt: $DIST(x,z) \leq DIST(x,y) + DIST(y,z)$.
Was passiert bei Übergang mittels kompatibler Transformation f:

- allgemein:
  $f(SIM(u,w)) \leq f(SIM(u,v)) + f(SIM(v,w))$.

  schwer interpretierbar:

- z.B. für $f(x) = 1 - x$:
  $SIM(u,v) + SIM(v,w) \leq 1 + SIM(u,w)$.

Anderer Zugang:

Invertierung der Dreiecksungleichung für Distanzen:

$$SIM(u,w) \geq SIM(u,v) + SIM(v,w).$$

Zusammen mit Reflexivität und Symmetrie:
$1 = SIM(u,u) \geq SIM(u,v) + SIM(v,u) = 2 \cdot SIM(u,v)$,
d.h. $SIM(u,v) \leq 0.5$ für beliebige u, v.
Analog:

$$SIM(u,w) \leq SIM(u,v) + SIM(v,w)$$

führt zu $SIM(u,v) \geq 0.5$ für beliebige u, v.

Spezielles Beispiel:

$$SIM([x_1, x_2], [y_1, y_2]) := Min(\{ 1 \ , \ card(\{ i \mid x_i = y_i\} \ \}. \tag{5}$$

mit
$SIM([0,0],[0,1]) = SIM([0,1],[1,1]) = SIM([0,0],[0,0]) = 1$
$SIM([0,0],[1,1]) = SIM([1,1],[0,0]) = 0$
daraus folgt:
$SIM([0,0],[0,1]) + SIM([0,1],[1,1]) > SIM([0,0],[1,1])$
$SIM([0,0],[1,1]) + SIM([1,1],[0,0]) < SIM([0,0],[0,0])$.
Kompatible Transformation f mit $f(0) = 1$ und $f(1) = 0$
für Übergang von SIM zu $DIST$ ergibt:

$$DIST([0,0],[0,1]) + DIST([0,1],[1,1]) < DIST([0,0],[1,1]).$$

$\Rightarrow$ Es gibt aus Ähnlichkeitsmaßen abgeleitete intuitive Distanzmaße, die die Dreiecksungleichung nicht erfüllen (insbesondere keine Metriken sind).

Bemerkung:
Es gibt stets relational kompatible Metrik für symmetrisches Ähnlichkeitsmaß $SIM : U \times U \rightarrow [0,1]$:

$$DIST(u,v) := \begin{cases} 0 & , \quad falls \quad u = v \\ \frac{1}{2} \cdot (2 - SIM(u,v)) & , \quad sonst. \end{cases} \tag{6}$$

(Dreiecksungleichung gilt wegen $\frac{1}{2} \leq DIST(u,v) \leq 1$.)

Kompositorische Maße

Das Lokal-Global-Prinzip

Gegeben: Attribute $A_1, ..., A_n$ für die Komponenten.

Definition

Lokal-Global-Prinzip für Ähnlichkeiten (bzw. Distanzen):

- *Lokales Ähnlichkeitsmaß* für Attribute A_i: $sim_i : \mathcal{R} \times \mathcal{R} \to \mathcal{R}$
- *Kompositionsfunktion* $COMP : \mathcal{R} \times \ldots \times \mathcal{R} \to \mathcal{R}$

Ein („globales") Ähnlichkeitsmaß $SIM : \mathcal{R} \times \ldots \times \mathcal{R} \to \mathcal{R}$ heißt *kompositorisch*, wenn es mit lokalen Ähnlichkeitsmaßen und einer Kompositionsfunktion definierbar ist:

$$SIM([q_1, ..., q_n], [u_1, ..., u_n]) = COMP(sim_1(q_1, u_1), ..., sim_n(q_n, u_n)).$$

Analog für Distanzen mit lokalen Distanzen $dist_i : \mathcal{R} \times \mathcal{R} \to \mathcal{R}$:

$$DIST([q_1, ..., q_n], [u_1, ..., u_n]) = COMP(dist_1(q_1, u_1), ..., dist_n(q_n, u_n)).$$

Beispiele:

- Gewichtete Hamming-Maße:

$$SIM([q_1, ..., q_n], [u_1, ..., u_n]) = \sum_{i=1,...,n} w_i \cdot sim_i(q_i, u_i). \tag{7}$$

- Euklidische Distanz:

$$DIST([q_1, ..., q_n], [u_1, ..., u_n]) = \sqrt{\sum_{i=1,...,n} dist_i(q_i, u_i)}. \tag{8}$$

  mit $dist_i(q_i, u_i) = |q_i - u_i|^2$.

- Fuzzy-Operatoren (t-norm, t-co-norm) als Kompositionsfunktionen mit linguistischen Termen als lokale Ähnlichkeitsmaße (s.u.).

Das Globale Monotonie-Axiom

$$\boxed{SIM(u,v) > SIM(u,w) \rightarrow \exists i \in \{1,...,n\} : sim_i(u_i,v_i) > sim_i(u_i,w_i).}$$

Gilt aber nicht immer. Gegenbeispiel („XOR-problem“):
$U := \{[0,0],[0,1],[1,0],[1,1]\}$

$SIM([u_1,u_2],[v_1,v_2]) = 1$ gdw. $XOR(u_1,u_2) = XOR(v_1,v_2)$, d.h.:

$$\begin{aligned} SIM([0,0],[1,1]) &= SIM([0,1],[1,0]) = 1 \\ SIM([0,0],[0,1]) &= SIM([0,0],[1,0]) \\ = SIM([1,1],[1,0]) &= SIM([1,1],[0,1]) = 0. \end{aligned}$$

folglich $SIM([0,0],[1,1]) > SIM([0,0],[0,1]$.

aber: $sim_1(0,1) \leq sim_1(0,0) = 1$ (Reflexivität) und $sim_2(0,1) = sim_2(0,1)$.

Lokales Monotonie-Axiom:

$$|u_i - v_i| < |u'_i - v'_i| \;\Leftrightarrow\; DIST(u_i, v_i) < DIST(u'_i, v'_i)$$
$$|u_i - v_i| < |u'_i - v'_i| \;\Leftrightarrow\; SIM(u_i, v_i) > SIM(u'_i, v'_i)$$

Gilt nicht immer, Gegenbeispiel:

$$sim(x, y) := \begin{cases} 1 & , \quad if \quad x = y \quad oder \quad x, y \in \mathcal{N} \\ 0 & , \quad otherwise. \end{cases} \tag{9}$$

Transformationen für Kompositorische Maße

Relational kompatible Transformationen auf der

- Lokalen Ebene: sim_i, $dist_i$
- Globalen Ebene: SIM, $DIST$

Beispiel für kompatible Transformationsfunktion:

$$\frac{1}{1+x}$$

Vgl. Beispiel (3):
Lokale Distanzen: $dist_i(x_i, y_i) = |x_i - y_i|$
Transformierte lokale Ähnlichkeitsmaße: $sim_i(x_i, y_i) = \frac{1}{1+|x_i - y_i|}$
$\Rightarrow$ Globales Ähnlichkeitsmaß SIM:

$$SIM([x_1, x_2, x_3], [y_1, y_2, y_3]) := \frac{1}{3} \cdot \sum_{i=1,2,3} \frac{1}{1 + |x_i - y_i|}.$$

Dagegen bei Transformation auf der globalen Ebene:

$$SIM'([x_1, x_2, x_3], [y_1, y_2, y_3]) := \frac{1}{1 + DIST([x_1, x_2, x_3], [y_1, y_2, y_3])} \tag{10}$$

d.h.:

$$SIM'([x_1, x_2, x_3], [y_1, y_2, y_3]) = \frac{1}{1 + \Sigma_{i=1,2,3} |x_i - y_i|}. \tag{11}$$

Komposition und Transformation nicht vertauschbar.

Lokale Maße, Merkmalstypen, Skalen

Merkmale (jetzt auch nicht-numerische Wertebereiche):

Qualitative Merkmale :

Beispiele:

- rot, grün, blau, ...
- ledig, verheiratet, geschieden, ...
- kalt, lau, warm, ...

Evtl. (halb-)geordnet, evtl. mit Ähnlichkeit.

Quantitative Merkmale :

Numerisch.
Geordnet.

Skalen:

Nominalskala : ungeordnet
Spezialfall *binäre Skala*: 2 Werte, z.B. „ja“, „nein“.

Ordinalskala/Rangskala : geordnet.

Metrische Skala : metrisch geordnet.
(Metrik: nicht-negativ, starke Reflexivität, Symmetrie, Dreiecksungleichung)

Ähnlichkeit z.B.

- definierbar durch Vergleichsrelation $V3$
  bei gegebener (Halb-)Ordnung ORD:

  $$\textit{y ist ähnlicher zu x als z} \;\; V3_{SIM}(x,y,z)$$
  $$:= (ORD(x,y) \wedge ORD(y,z)) \vee (ORD(z,y) \wedge ORD(y,x))$$

- definierbar als inverse Distanzen

- gegeben durch Synonyme, ähnliche Begriffe, ...

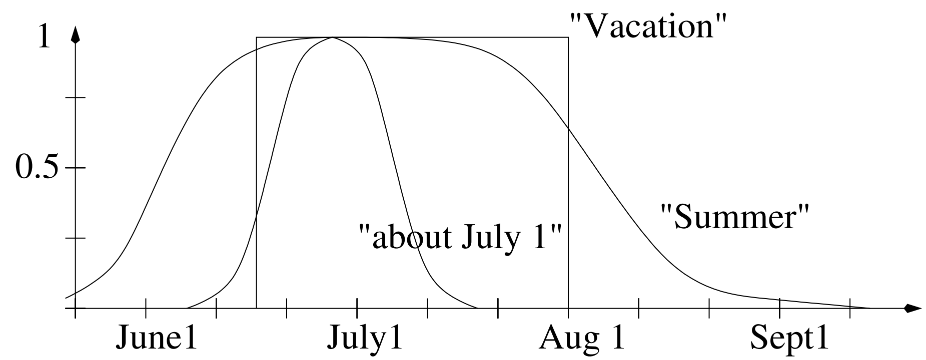
- gegeben durch unscharfe Begriffe (linguistische Terme)

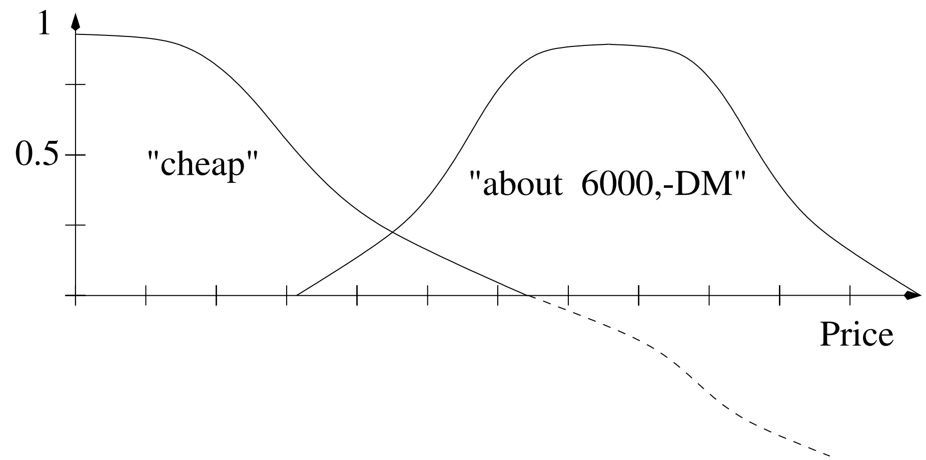
Ähnlichkeit zwischen konkreten Werten und linguistischen Termen

- Unscharfe Begriffe:
  billig
  ungefähr 1000 DM
  Sommer

- Scharfe Begriffe:
  Ferienzeit

- Ähnliche Begriffe:
  Sommer – Ferienzeit
  6000 DM – 5499 DM





Wort(-gruppen)-Ähnlichkeit in Texten

- Identische Bedeutung
  - Beugung, Zahl (`Printers`)
  - Sprache (`Drucker`)
  - Synonyme (`LaserJet`)

- Ähnliche Bedeutung
  - ähnlich (`Plotter`)
  - Wortgruppe (`Output Device`)
  - Gemeinsames Auftreten (`PostScript`)

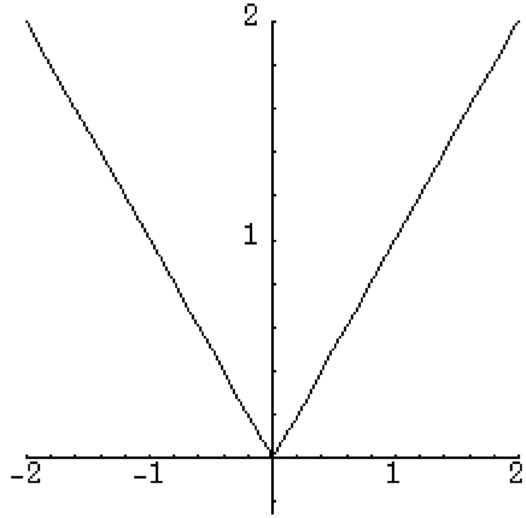
$\Rightarrow$ Kontext-abhängig

Lokale Transformation: Ähnlichkeit $\leftrightarrow$ Distanz

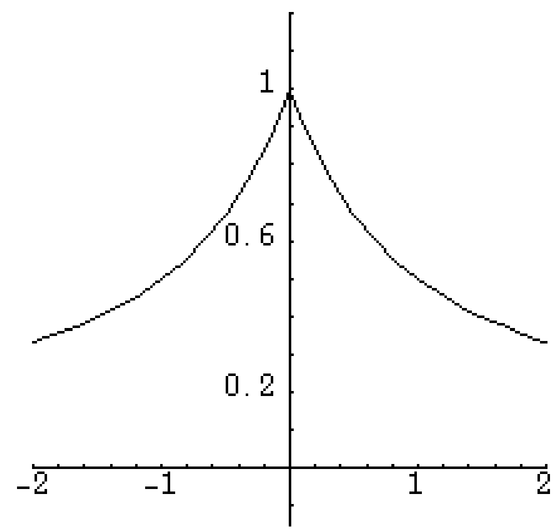
Kompatibilität: $SIM(x,y) \geq SIM(u,v) \Leftrightarrow DIST(x,y) \leq DIST(u,v)$

Transformationsfunktion z.B.

$$\frac{1}{1+x}$$



$dist_i(x_i, y_i) = |x_i - y_i|$



$sim_i(x_i, y_i) = \frac{1}{1+|x_i - y_i|}$

Kombination lokaler Ähnlichkeiten

Lokal-Global-Prinzip für Ähnlichkeiten (bzw. Distanzen)
bei beliebigen Wertebereichen W_i der Attribute A_i.

- *Lokales Ähnlichkeitsmaß* für Attribute A_i: $sim_i : W_i \times W_i \to \mathcal{R}$
- *Lokale Distanz* für Attribute A_i: $dist_i : W_i \times W_i \to \mathcal{R}$
- *Kompositionsfunktion* $COMP : \mathcal{R} \times \ldots \times \mathcal{R} \to \mathcal{R}$

Kompositorisches Ähnlichkeitsmaß:

$$SIM([q_1, ..., q_n], [u_1, ..., u_n]) = COMP(sim_1(q_1, u_1), ..., sim_n(q_n, u_n)).$$

Kompositorische Distanz:

$$DIST([q_1, ..., q_n], [u_1, ..., u_n]) = COMP(dist_1(q_1, u_1), ..., dist_n(q_n, u_n)).$$

Beispiele für **reelle Merkmalswerte**:

Euklidischer Abstand:

$$DIST(x,y) := \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

Minkowski-Metriken

$$DIST(x,y)^{(r)} := (\sum_{i=1}^{n} |x_i - y_i|^r)^{\frac{1}{r}}$$

für $r = 2$: euklidische Metrik
für $r = 1$: Manhattan-Metrik
für $r \to \infty$: Maximum-Norm

$$DIST(x,y)^{(\infty)} = Max\{|x_i - y_i| \ / \ i = 1, ..., n\}$$

Fuzzy-Theorie:
t-norm, z.B. Minimum (verallgemeinerte "Konjunktion")
t-co-norm, z.B. Maximum (verallgemeinerte "Alternative")

Probleme:

- Skalierung (Dehnung/Stauchung)
  $\Rightarrow$ Normierung aller Merkmale z.B. auf Skala $[0,1]$ durch merkmalsabhängigen Faktor σ_i, z.B.:

$$DIST(x,y) := \sqrt{\sum_{i=1}^{n} (\sigma_i \cdot |x_i - y_i|)^2}$$

- Wichtung von Merkmalen durch merkmalsabhängiges Gewicht γ_i, z.B.:

$$DIST(x,y) := \sqrt{\sum_{i=1}^{n} \gamma_i \cdot |x_i - y_i|^2}$$

Skalierung durch Wichtung i.a. nicht ersetzbar.

Ausgleich bei fehlenden Werten (variablem n):

$$dist(x,y) := \sqrt{\frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|^2}$$

(andere Varianten: optimistisch/pessimistisch)

Beispiele für **Binäre Merkmalswerte**: $x = [x_1, ..., x_n]$, $y = [y_1, ..., y_n]$ jeweils aus $\{0, 1\}^n$

Kontingenztafel:

$$\begin{aligned}
a &= \sum_{i=1}^{n} x_i \cdot y_i \\
b &= \sum_{i=1}^{n} x_i \cdot (1 - y_i) \\
c &= \sum_{i=1}^{n} (1 - x_i) \cdot y_i \\
d &= \sum_{i=1}^{n} (1 - x_i) \cdot (1 - y_i)
\end{aligned}$$

Hamming-Abstand: $DIST(x, y) = b + c$

Simple-Matching-Coefficient: $SIM(x, y) = \frac{a+d}{a+b+c+d}$

Gewichtete Strategien („optimistisch“: $g > 0,5$ bzw. „pessimistisch“: $g < 0,5$):

$$SIM(x, y) = \frac{g \cdot (a + d)}{g \cdot (a + d) + (1 - g) \cdot (b + c)}$$

Reduktion von nominalen Skalen auf binäre Skalen:

Bsp.: Merkmal $Farbe \in \{rot, grün, blau\}$
übersetzen in binären Merkmalsvektor $[w_{rot}, w_{grün}, w_{blau}]$ mit $w_{rot} = 1 \Leftrightarrow Farbe = rot$ usw.

Dabei ein bzgl. 0/1 asymmetrisches Maß benutzen, z.B.

$$SIM(x, y) = \frac{a}{a + b + c}$$

Contrast Rule (Tversky):

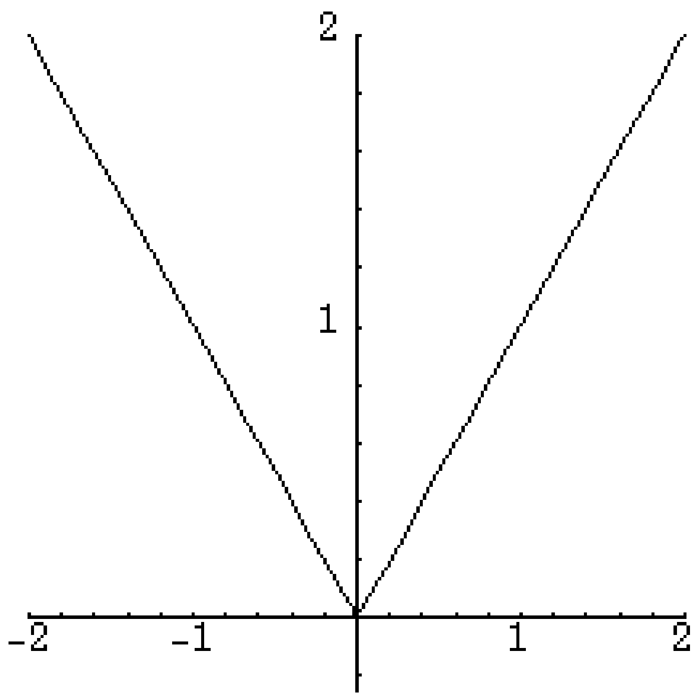
Objekte (S=„Situation“, P=„Prototyp“) beschrieben als Menge ihrer Eigenschaften.
Jede Eigenschaft wird mit einem Gewicht g_i bewertet.

$$SIM(S, P) = \alpha \cdot \sum_{i \in S \cap P} g_i \quad - \quad \beta \cdot \sum_{i \in P - S} g_i \quad - \quad \gamma \cdot \sum_{i \in S - P} g_i$$

(α, β, γ als weitere Gewichte)

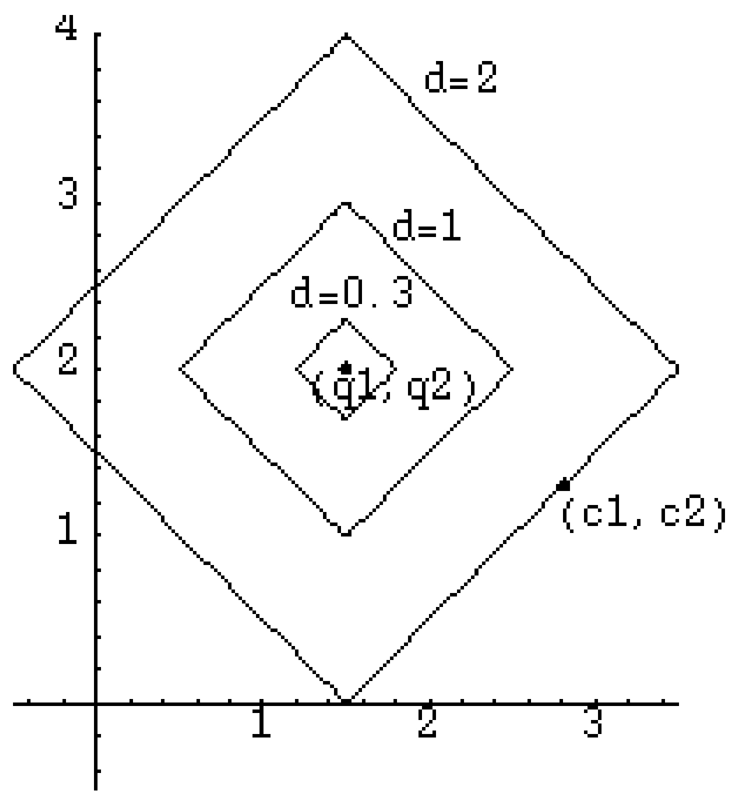
Unterschiedliche Charakteristiken (=Linien gleicher Ähnlichkeit/Distanz)

Lokale Distanzen

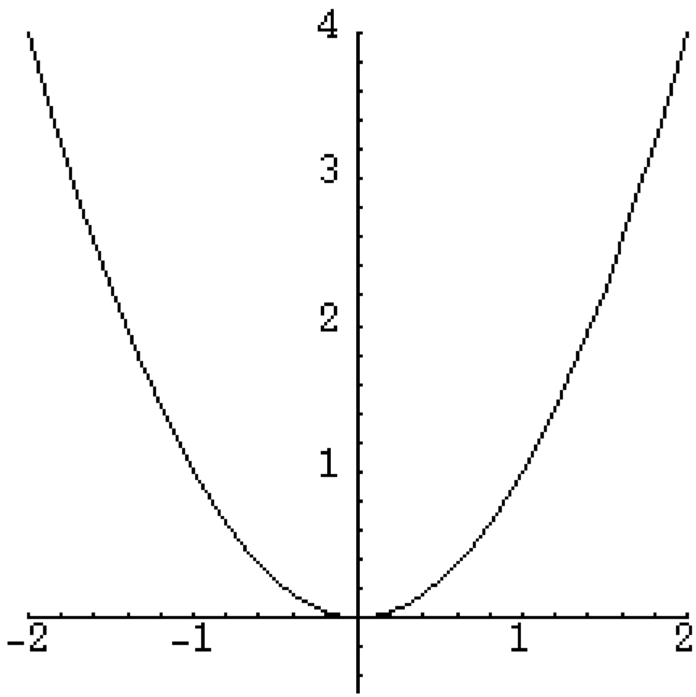


$f(x) = |x|$

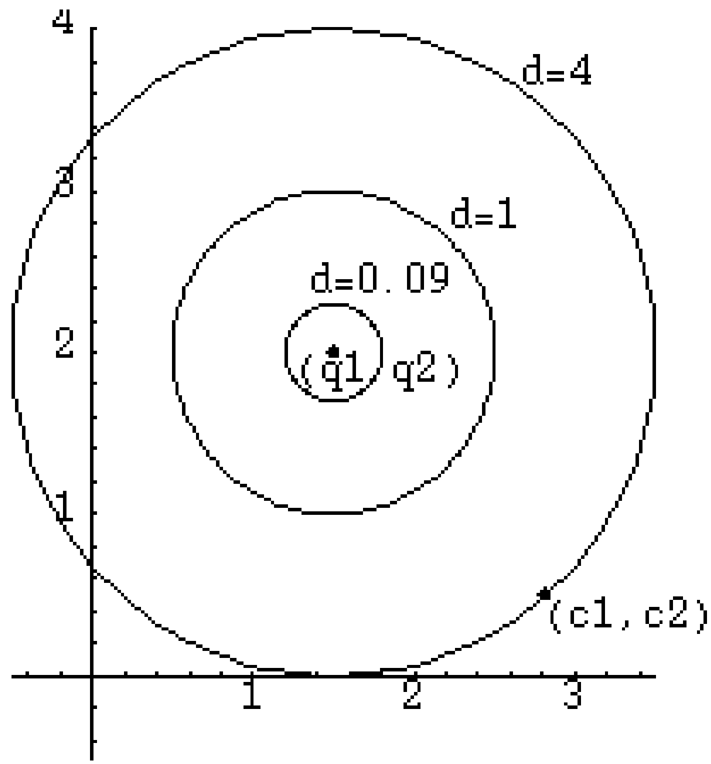
Globale Distanzen



$|q_1 - c_1| + |q_2 - c_2| = d$

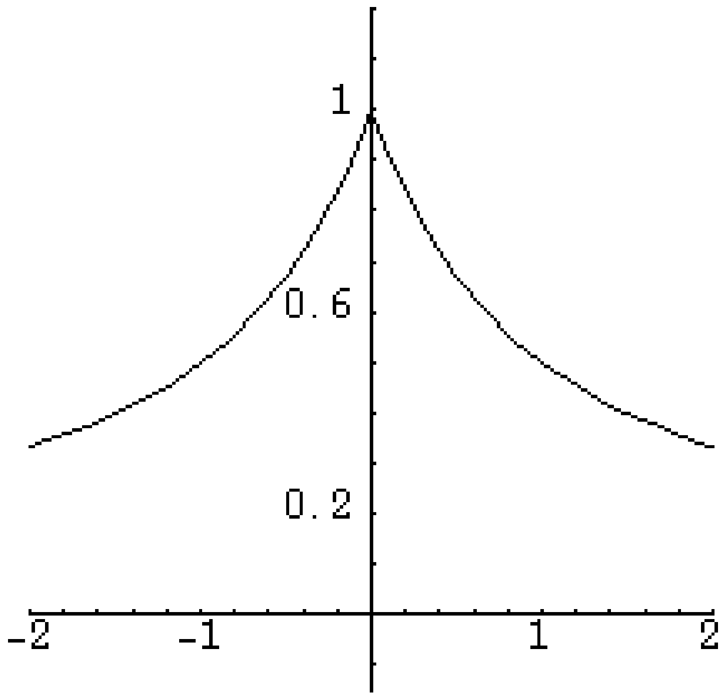


$f(x) = x^2$



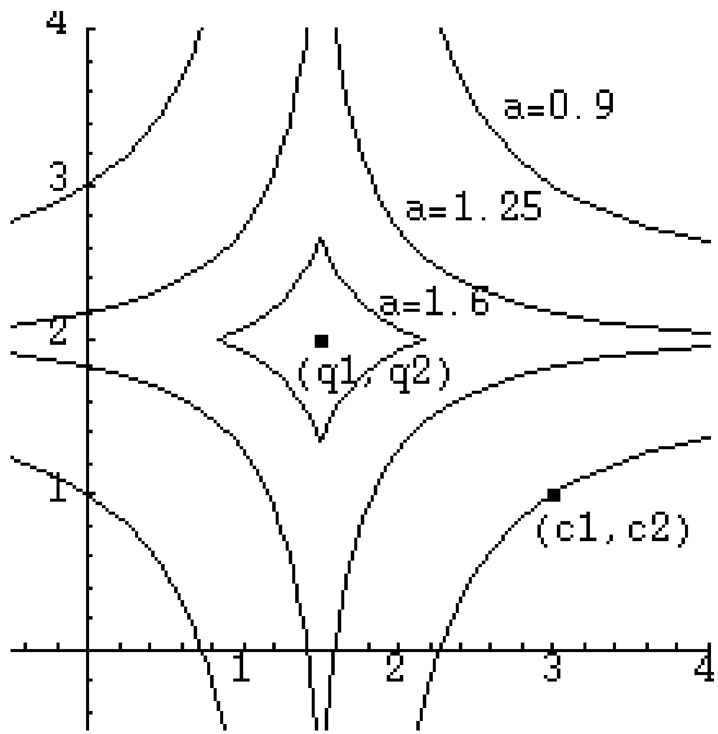
$(q_1 - c_1)^2 + (q_2 - c_2)^2 = d$

Lokale Ähnlichkeiten

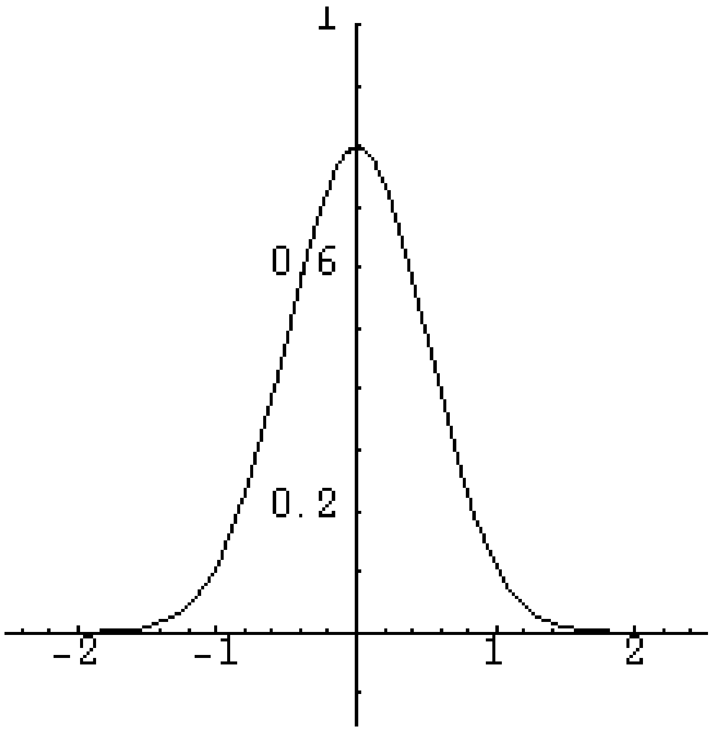


$f(x) = 1/(1 + |x|)$

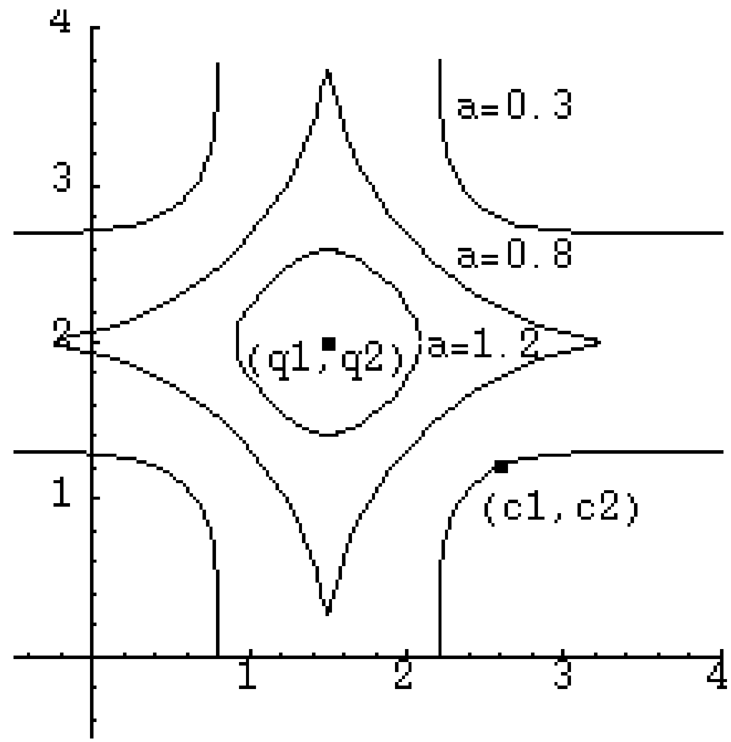
Globale Ähnlichkeiten



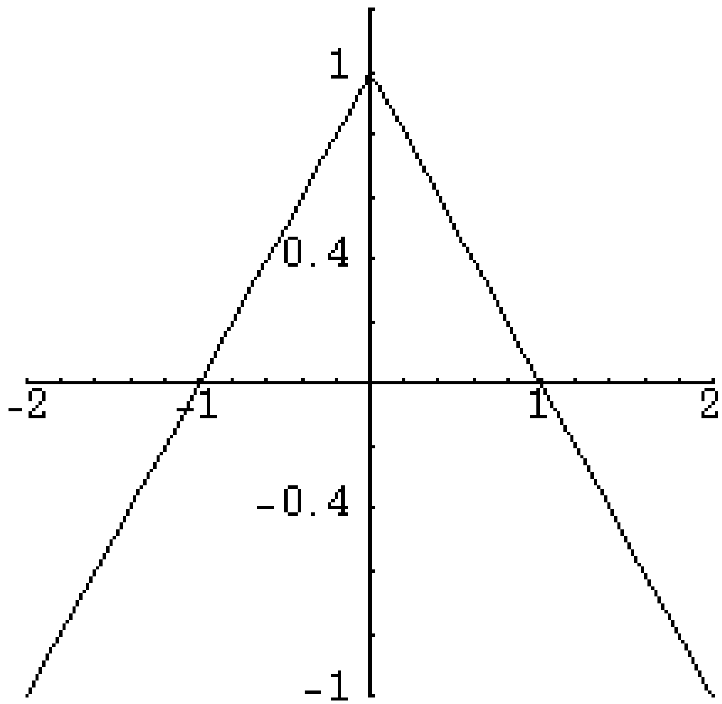
$1/(1 + |q_1 - c_1|) + 1/(1 + |q_2 - c_2|) = a$

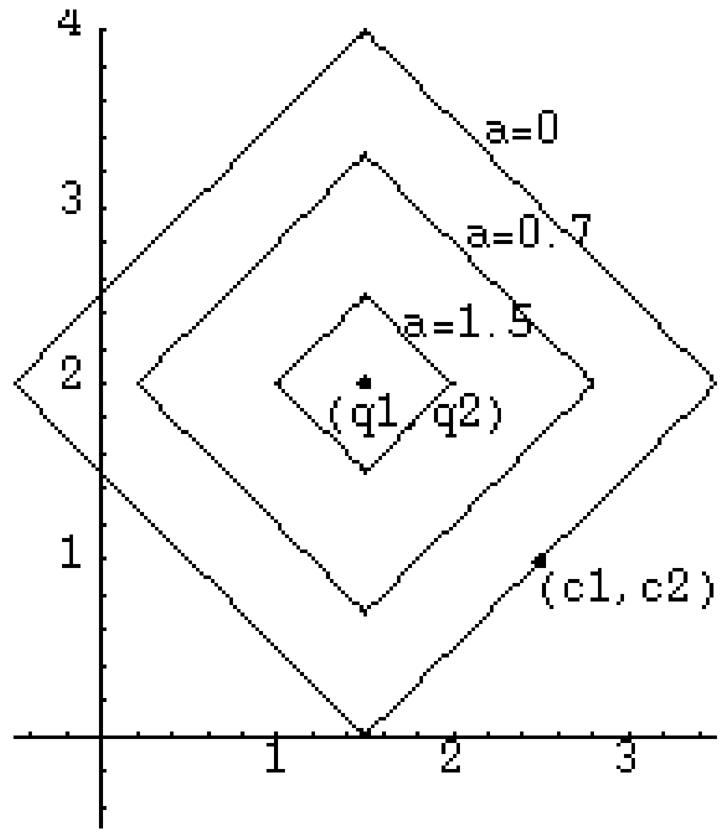


$f(x) = Gauss(x, 0, 0.5)$



$Gauss(q_1 - c_1, 0, 0.5) + Gauss(q_2 - c_2, 0, 0.5) = a$





$f(x) = 1 - |x|$

$(1 - |q_1 - c_1|) + (1 - |q_2 - c_2|) = a$

Intuitive Bedeutung kompositorischer Maße

Speziell *Gewichtete Summen*:
Akkumulation lokaler Werte zu globalem Wert mit (fixierten) Gewichten der Attribute

- Distanz:
  akkumuliert *Argumente für Ausschluß*

  $\boxed{\text{Akkumulation für Zurückweisung}}$

- Ähnlichkeit:
  akkumuliert *Argumente für Annahme*

  $\boxed{\text{Akkumulation für Akzeptanz}}$

Akkumulation für Zurückweisung :

Einzelne große Distanz → Zurückweisug.
Nächster Nachbar: *Alle* lokalen Werte sind nah.

Methode: $\boxed{\text{Entscheidungsbaum}}$
("top down pruning")

Akkumulation für Akzeptanz :

Hinreichend viele lokale Werte ähnlich → Akzeptanz.

→ Kompromisse (!)

Methode: $\boxed{\text{Case Retrieval Netze}}$
("bottom up collecting")

Unterschiede Ähnlichkeit/Distanz:

- Ähnlichkeit akkumuliert Argumente für Akzeptanz.
- Distanz akkumuliert Argumente für Zurückweisung.

Grundproblem bezüglich Ähnlichkeit/Distanz:

1. Ähnlichkeit/Distanz impliziert Nützlichkeit.
2. Ähnlichkeit/Distanz basiert auf a-priori-Fakten.
3. Ähnlichkeit/Distanz als quantitatives Maß.

Dafür

- Auswahl geeigneter Merkmale.
- Auswahl geeigneter (intuitiver) Maße für Ähnlichkeit/Distanz:
    1. Lokale Ähnlichkeiten/Distanzen
    2. Kompositionsfunktion
    3. Akkumulationsprinzip (Akzeptanz vs. Zurückweisung)

Globale/lokale Monotonie unterstützen Intuitivität.