

Merkmalsraum

Foundations • The Valuation Engineer

Bert Craytor • 1. August 2026

Version 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Der Merkmalsraum einer Klasse heterogener Güter ist der Vektorraum, dessen Koordinaten die von Käufern bewerteten Attribute indizieren und in dem jedes Gut durch den Vektor seiner Attributausprägungen dargestellt wird.

Intuitive Einordnung. Der vorangehende Eintrag hat festgehalten, dass ein heterogenes Gut ein eigenes Bündel von Merkmalen ist. Die natürliche Anschlussfrage lautet: Wie sollen wir solche Bündel formal beschreiben und vergleichen? Die Antwort, die auf Lancaster (1966) zurückgeht, besteht darin, jedes Gut als Punkt in einem Vektorraum zu behandeln, dessen Achsen die Merkmale sind, auf die Käufer Wert legen.

In dieser Darstellung ist eine Wohnimmobilie in Pacifica nicht mehr „ein Haus in Park Pacifica“, sondern ein Vektor mit Koordinaten für Wohnfläche (GLA), Grundstücksgröße, Aussicht, Zustand, Schulbezirk, Entfernung zum Ozean und so weiter. Zwei Immobilien, die alle diese Attributwerte teilen, würden denselben Punkt im Merkmalsraum einnehmen; Immobilien, die sich in irgendeinem Attribut unterscheiden, besetzen verschiedene Punkte. Die Geometrie des Raums kodiert, wie nah oder wie verschieden zwei Immobilien sind — und zwar so, dass alle Dimensionen erfasst werden, in denen Käufer Unterscheidungen treffen.

Die Dimensionalität des Raums hängt davon ab, welche Attribute für den untersuchten Markt relevant sind. Für ein typisches Wohnimmobiliengutachten mag der relevante Raum ein Dutzend Dimensionen haben: physische (GLA, Grundstück, Schlafzimmer, Bäder, Alter, Qualität, Zustand), lagebezogene (Schulbezirk, Aussicht, Nähe zu Annehmlichkeiten oder Störquellen) und rechtlich-ökonomische (Bauplanungsrecht, Dienstbarkeiten). Für ein Gewerbegutachten verschieben sich die relevanten Dimensionen zu Cashflow-Attributen (Mietvertragsstruktur, Mieterbonität, Belegung), aber die zugrunde liegende geometrische Idee ist dieselbe.

Was den Merkmalsraum nützlich macht, ist, dass er eine qualitative Praktikerintuition („diese Comps sind dem Bewertungsobjekt ähnlich“) in eine quantitative Behauptung verwandelt („diese Comps liegen dem Bewertungsobjekt im Merkmalsraum nahe“). Sobald Distanz definiert ist, lässt sich jede Entscheidung im Vergleichswertverfahren als Aussage über Lage im Raum untersuchen.

Wo Gutachter dem Konzept begegnen. Gutachter arbeiten ständig im Merkmalsraum, meist ohne ihn so zu nennen. Jede Zeile eines Vergleichsrasters ist eine Koordinatenachse; jeder Comp ist ein Zeilenvektor von Koordinatenwerten; jede Anpassung ist der Versuch, sich entlang einer Achse zu bewegen und die anderen festzuhalten.

Die acht Pacifica-Vergleichsobjekte, die wir in dieser Ausgabe durchgehend verwenden, illustrieren die Konstruktion direkt. Jeder Comp ist ein Punkt in einem vierdimensionalen Merkmalsraum, aufgespannt von GLA, Grundstücksgröße, Aussicht und Zustand:

Comp	GLA (sf)	Grundstück (sf)	Aussicht	Zustand	Verkaufspreis
A	1,650	9,000	ja	gut	\$1,425,000
B	1,650	7,200	nein	gut	\$1,280,000
C	1,850	7,500	nein	gut	\$1,407,500
D	1,450	7,000	nein	gut	\$1,155,000
E	1,700	8,000	ja	ausgezeichnet	\$1,470,000
F	1,550	7,100	nein	durchschnittl.	\$1,177,500
G	1,700	10,500	nein	gut	\$1,392,500
H	1,600	7,300	nein	gut	\$1,312,500

Comp A und Comp B aus dem vorangehenden Eintrag sind die ersten beiden Zeilen; die übrigen sechs erweitern das Beispiel auf Weisen, die wir im Rest der Ausgabe ausnutzen werden.

Jeder Comp ist ein Punkt in $\mathbb{R}^4$:

- Comp A liegt bei (1650, 9000, 1, 2).
- Comp B liegt bei (1650, 7200, 0, 2).
- Comp G liegt bei (1700, 10500, 0, 2).

Die Koordinaten für Aussicht und Zustand sind numerisch kodiert (Aussicht als $\{0, 1\}$, Zustand als $\{1, 2, 3\}$); der Preis ist *keine* Koordinate des Merkmalsraums, sondern eine Funktion der Position darin.

Warum es für die Vertretbarkeit wichtig ist. Das Merkmalsraum-Rahmenwerk schärft, was Comp-Auswahl eigentlich ist. Ein Vergleichsobjekt ist nicht bloß „eine ähnliche Immobilie“, sondern ein Punkt in der relevanten Region des Merkmalsraums nahe dem Bewertungsobjekt. Je näher ein Comp dem Bewertungsobjekt im Raum liegt, desto kleiner sind die erforderlichen Anpassungen und desto verlässlicher wird der angezeigte Wert.

Diese Umformung legt mehrere Praxisfragen offen, die das informelle Raster verdeckt:

- *Welche Dimensionen zählen?* Die Wahl der Achsen ist eine Modellierungsentscheidung. „Aussicht“ als Achse aufzunehmen ist nur sinnvoll, wenn der Markt Aussicht eigenständig bepreist; eine „Feng-Shui-Ausrichtung“ aufzunehmen ist nur sinnvoll, wenn ein Marktsegment sie bepreist. Vertretbare Bewertung erfordert, die Achsen zu verteidigen, nicht nur die Werte entlang der Achsen.
- *Welches Distanzmaß gilt?* Zwei Comps können sich um \$200,000 im Preis unterscheiden, weil sie im Merkmalsraum weit auseinanderliegen, aber die Distanzberechnung erfordert eine Skalierung für jede Achse. Untransformierte euklidische Distanz behandelt einen Quadratfuß GLA als äquivalent zu einem Quadratfuß Grundstück, was der Markt nicht tut.
- *Liegen die Comps in einer Region, in der sich die Preisfunktion gutartig verhält?* Sitzt das Bewertungsobjekt in einer dünn besetzten Ecke des Merkmalsraums, liegt kein Comp in seiner Nähe, und die Wertindikation ist notwendigerweise extrapolativ.

Durchgerechnetes Bewertungsbeispiel. Die acht Pacifica-Comps besetzen eine relativ kompakte Region des vierdimensionalen Merkmalsraums. Betrachten wir die Frage, welcher Comp einem hypothetischen Bewertungsobjekt bei (1,700, 7,500, 1, 2) — 1,700 sf, 7,500 sf Grundstück, Aussicht, Zustand „gut“ — am „nächsten“ liegt.

Verwenden wir naiv die euklidische Distanz mit unskalierten Achsen, dominiert die Grundstücksdimension, weil ihre Werte (7,000–10,500) numerisch viel größer sind als Aussicht (0–1) oder Zustand (1–3). Distanzen zum Bewertungsobjekt:

Comp	Distanz (unskaliert, euklidisch)	Anmerkungen
A	1,501	Aussicht passt; Grundstück dominiert
B	308	Aussicht differiert; Grundstück nah
C	230	Aussicht differiert; nach Rohdistanz am nächsten
D	559	Aussicht differiert; GLA abweichend
E	502	Aussicht passt; Zustand differiert
F	422	Aussicht und Zustand differieren
G	3,000	Aussicht differiert; Grundstück weit entfernt
H	224	Aussicht differiert; am nächsten von allen

Nach Rohdistanz ist Comp H der nächste Nachbar. Aber Comp H hat keine Aussicht, das Bewertungsobjekt schon. Das naive Maß hat uns in die Irre geführt: Die numerische Skala der Grundstücksdimension hat die kategoriale Aussichtsdimension verdrängt, die für den Preis tatsächlich mehr zählt.

Ein vertretbares Distanzmaß würde jede Achse mit ihrem impliziten Preis skalieren — grob gesagt damit, wie viele Dollar eine Einheit Variation entlang dieser Achse wert ist. Sind die Achsen erst nach impliziten Preisen skaliert, wird Aussicht (im Wert von ~\$100,000) nicht mehr vom Grundstück (~\$25 pro Quadratfuß) dominiert. Comp A, der beim Merkmal Aussicht mit dem Bewertungsobjekt übereinstimmt, tritt dann als das geeignete primäre Vergleichsobjekt hervor.

Genau das liefert die Maschinerie der impliziten Preise, die in den folgenden Einträgen entwickelt wird: eine prinzipiengeleitete Skalierung des Merkmalsraums, die „Nähe“ von einer geometrischen Vermutung in eine marktabgeleitete Berechnung verwandelt.

Literatur.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
 Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Querverweise: heterogenes Gut; hedonische Preisfunktion; impliziter Preis.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (002-characteristics-space.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 002-characteristics-space.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 002
%% Title: Characteristics Space
%% Author: Bert Craytor
%% Concept: characteristics_space
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('.././../././tools/dictionary/dictionary').
```

```

:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(characteristics_space, "Characteristics Space", economic_theory,
%      "The characteristics space of a class of heterogeneous goods is
%      the vector space whose coordinates index the attributes that
%      buyers value, and in which each good is represented by the
%      vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(heterogeneous_good, ...).
% term(hedonic_price_function, ...).
% term(implicit_price, ...).
%
% generalized_by(characteristics_space, lancaster_1966).
% depends_on(characteristics_space, heterogeneous_good).
% -----

:- use_terms([ characteristics_space, heterogeneous_good,
                hedonic_price_function, implicit_price,
                comp_selection, sales_comparison_approach,
                real_estate ]).
:- use_relations([related/2, depends_on/2, generalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(characteristics_space).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.002').

% =====
% LOCAL VOCABULARY
% =====
%
% Narrow article-local field names + the dimensions of the worked
% example's characteristics space.

local_term(dimension).
local_term(coordinate).
local_term(euclidean_distance).
local_term(distance_metric).
local_term(nearest_neighbor).

% =====
% FORMAL ATTRIBUTION
% =====

% The construction is due to Lancaster (1966), as noted in citations.
attributed_construction(characteristics_space, lancaster_1966).

% =====
% THE ARTICLE'S CLAIM ABOUT REPRESENTATION
% =====
%
% A heterogeneous good becomes a point in a vector space. Each axis is
% a characteristic buyers value; each good is the vector of its
% attribute levels.

```

```

representation_of(heterogeneous_good, point_in_vector_space).

% Examples of dimensions relevant to residential appraisal.
residential_dimension(gla_sqft).
residential_dimension(lot_sqft).
residential_dimension(view).
residential_dimension(cond).
residential_dimension(bedrooms).
residential_dimension(baths).
residential_dimension(age).
residential_dimension(quality).
residential_dimension(school_district).
residential_dimension(distance_to_amenity).
residential_dimension(zoning).

% Examples of dimensions relevant to commercial appraisal - included
% to make the article's point that the dimensions shift with market.
commercial_dimension(lease_structure).
commercial_dimension(tenant_quality).
commercial_dimension(occupancy).
commercial_dimension(cash_flow_profile).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% Mapping appraisal practice onto the characteristics-space framework.

practice_to_space_correspondence(
    sales_comparison_grid_line, coordinate_axis,
    "A grid line is an axis.").
practice_to_space_correspondence(
    comp_row, row_vector,
    "A comp is a vector of coordinate values.").
practice_to_space_correspondence(
    adjustment, movement_along_axis,
    "An adjustment moves along one axis holding others fixed.").

% =====
% WORKED EXAMPLE
% =====
%
% The 8 Pacifica comps are points in a 4D characteristics space spanned
% by GLA, lot_sqft, view, cond_numeric. Bring in the shared dataset.

:- consult('../..//kb/pacifica_comps').

% Axes of the chosen space for this article's analysis.
axis(gla_sqft).
axis(lot_sqft).
axis(view).
axis(cond_numeric).

% Numeric encoding for non-numeric axes.
encode(view, no, 0).
encode(view, yes, 1).
encode(cond_numeric, 1, 1).
encode(cond_numeric, 2, 2).
encode(cond_numeric, 3, 3).

```

```

%! coord(?Property, ?Axis, ?Numeric) is nondet.
%
% Numeric coordinate of Property along Axis. Handles symbolic attribute
% values by encoding them.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V ; encode(A, V, N) ).

%! point(?Property, -Vector) is det.
%
% Vector representation of Property in the 4D characteristics space.

point(P, [GLA, Lot, View, Cond]) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond).

% =====
% THE HYPOTHETICAL SUBJECT
% =====
%
% The article tests distance metrics against this hypothetical property.

hypothetical_subject(subject).
attribute_value(subject, gla_sqft, 1700).
attribute_value(subject, lot_sqft, 7500).
attribute_value(subject, view, yes).
attribute_value(subject, cond_numeric, 2).

% =====
% NAIVE EUCLIDEAN DISTANCE (the article's straw distance metric)
% =====
%
% Computes the unscaled Euclidean distance between Property and the
% hypothetical subject. Used to demonstrate that unscaled distances
% mislead: the lot dimension dominates because its numeric range is
% much larger than view (0..1) or cond (1..3).

naive_distance_to_subject(P, D) :-
    point(P, V_P),
    point(subject, V_S),
    sum_of_squares(V_P, V_S, S),
    D is sqrt(S).

sum_of_squares([], [], 0).
sum_of_squares([X|Xs], [Y|Ys], S) :-
    sum_of_squares(Xs, Ys, S0),
    D is X - Y,
    S is S0 + D * D.

% =====
% THE ARTICLE'S CENTRAL OBSERVATION
% =====
%
% Under naive Euclidean distance, Comp H is the nearest neighbor. But
% Comp H has no view and the subject does; view matters more to price

```

```

% than the naive metric admits. The fix (developed in entries 003-005)
% is to scale each axis by its implicit price.

claim(naive_metric_misleads,
    "Unscaled Euclidean distance crowds out small-range axes (like view) under large-range axes (
    ↪ like lot_sqft). A market-defensible distance scales each axis by its implicit price.>").

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====
%
% Three questions the characteristics-space framing makes explicit.

defensibility_question(which_dimensions,
    "The choice of axes is a modeling decision. Defending an appraisal requires defending the axes,
    ↪ not just the values on them.>").
defensibility_question(distance_metric,
    "Untransformed Euclidean distance treats one sqft of GLA as equivalent to one sqft of lot,
    ↪ which the market does not. Distance requires a scaling per axis.>").
defensibility_question(extrapolation_region,
    "If the subject sits in a sparsely-sampled corner of the space, no comp is truly close and the
    ↪ indication is extrapolative.>").

% =====
% INFERENCE RULES
% =====

%! same_point(?A, ?B) is nondet.
%
% Two properties occupy the same point in the characteristics space.
% (Doesn't occur in the running example, but the article notes that it
% would if all coordinates matched.)

same_point(A, B) :-
    point(A, V),
    point(B, V),
    A \== B.

%! nearest_neighbor_by_naive_distance(-P) is semidet.
%
% The comp closest to the subject under the naive Euclidean metric.
% Used to illustrate the article's point that the naive metric chooses
% Comp H, not Comp A.

nearest_neighbor_by_naive_distance(P) :-
    findall(D-Q,
        ( property(Q), Q \== subject, naive_distance_to_subject(Q, D) ),
        Pairs),
    keysort(Pairs, [_-P|_]).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```