

スパース推定を用いた変数選択と ヘドニック・アプローチによる不動産価格形成要因の分析

**Analysis of Real Estate Price Formation Factors Using Variable Selection and
Hedonic Approach Using Sparse Estimation**

中島 健希 (Tatsuki Nakashima)
u120031@st.pu-toyama.ac.jp

富山県立大学 工学部 情報システム工学科 情報基盤工学講座

11:10-11:25, Monday, February 10, 2025
N210, Toyama Prefectural University

1. はじめに

3. ヘドニック・ア
プローチによる不
動産価格形成要因
の分析

4. 提案手法

5. 数値実験並び
に考察

6. おわりに

1.1 本研究の背景

2/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

■ 既存研究

- 公示地価を対象にヘドニック・アプローチを実施
- 候補変数が少ない

■ ヘドニック・アプローチの概要と課題

- 物品・サービスの価格を特性ごとに分解し、市場価値を評価
- 多重共線性：説明変数の高相関が推定を不安定に
- 欠落変数バイアス：要因の見落としによる誤推定
- 交互作用・非線形効果を十分に考慮できていない

1.2 本研究の目的

3/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

- 不動産の実勢価格を対象に分析
- 一次項, 二次項, 交互作用項を含め多数の変数を候補に
- AEN (Adaptive Elastic Net) を用いて説明変数の絞り込みを行う.
- AEN は, $L1 \cdot L2$ 正則化を組み合わせることで, 多重共線性に対応しながら重要変数を選択する手法である.
- 適応的な正則化パラメータを導入することで, 非線形性や交互作用効果を考慮することが可能である.
- AEN を使用した後, 不動産価格形成要因の分析を行う.

1.2 本研究の目的

4/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

不動産とは土地と建物

誰でもわかる
不動産売買

不動産



2.1 多様な要因を考慮したデータセットの作成

5/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

サイバー空間からの多様なデータの取得

- 空き地の面積や、金融機関までの距離など、地理的な特徴を表すデータを、地図画像や地図情報サイトから取得する。

2.1 多様な要因を考慮したデータセットの作成

6/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

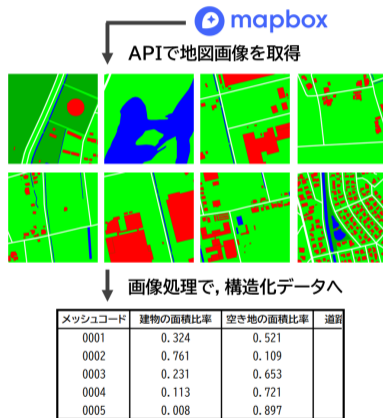


図11 地図画像にもとづく説明変数の抽出



スクレイピング、ジオコーディング

コンビニエンスストア	金融機関	レジャー施設	
[137.203, 36.603]	[136.310, 36.321]	[137.433, 36.692]	[137.4
[137.214, 36.651]	[137.002, 36.621]	[137.562, 36.832]	[137.5
[137.195, 36.599]	[137.322, 36.691]	[137.203, 36.293]	[137.2
[137.201, 36.709]	[136.893, 36.761]	[137.543, 36.654]	[137.5
[137.195, 36.541]	[137.312, 36.782]	[137.143, 36.932]	[137.1

図12 地図情報サイトからのスクレイピング

2.2 データセットの前処理

7/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

データセットの前処理

- 土地価格の予測モデルを構築するためには、多くの説明変数を考慮する必要がある。
- 生データには欠損値や外れ値が含まれていることが多く、モデルの精度を低下させる恐れがある。
- データセットの内容を把握し、必要な前処理を行うことが不可欠。

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

ステップ	前処理手法	目的
ステップ1	欠損値の処理	欠損データを処理する (削除、平均値または中央値で補完など)
ステップ2	外れ値の処理	外れ値を検出し、除去または補正する
ステップ3	カテゴリカル変数のダミー変数化	カテゴリ変数を数値データに変換 (One-Hotエンコーディングなど)
ステップ4	二次項の作成	各変数の二次項 (平方項) を作成し、非線形効果を考慮する
ステップ5	交互作用項の作成	複数の変数間の交互作用項を作成し、相互作用効果を評価する
ステップ6	Zスコアによる標準化	特徴量を平均0、標準偏差1にスケールリングし、異なるスケールの変数を揃えることができる。

2.3 説明変数の選定手法

9/27

説明変数の選定手法

説明変数選定は、回帰モデルや機械学習モデルの性能を最大化するために欠かせない工程である。以下のポイントが重要である

- **モデルの精度向上:** 適切な説明変数を選ぶことで、モデルの予測精度が向上し、過剰適合（オーバーフィッティング）を防ぐことができる。
- **データ理解の深化:** 説明変数の選定はデータの理解を深めるプロセスであり、モデルの解釈性を高めることができる。
- **計算効率の向上:** 不要な変数を排除することにより、計算効率が高まり、モデルが解釈しやすくなる。
- **適切な手法選択:** 各手法はデータや目的に応じて使い分けるべきであり、変数選定が適切に行われることが良い予測モデルを作成するための鍵である。

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

2.3 説明変数の選定手法

10/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

手法	詳細
フィルタ法	相関関係や統計的有意性に基づいて変数を選定する。相関係数やt検定、F検定を用いて説明変数間の関係を評価。相関係数が高い場合は多重共線性の問題がある可能性があり、削除が推奨される。
ラッソ回帰	回帰係数にL1ノルム（絶対値の合計）をペナルティとして加え、冗長な変数を自動的に排除する。過剰適合を防ぎ、簡素なモデルを得る。
ステップワイズ法	変数選定を段階的に行う方法。前進選択法、後退消去法、双方向法があり、AICやBICを基に最適なモデルを選定。
主成分分析	説明変数間の相関を低減し、次元を削減する方法。PCAを用いることで、相関がなくなり、過剰適合を防ぎつつ計算効率を向上させる。
共分散選択基準	VIFや条件数を用いて多重共線性を評価。多重共線性がある場合、変数を削除または標準化する。

3.1 ヘドニック法とその問題点

11/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

ヘドニック法とは

- 製品の価格に対する特性の影響を定量化する手法.
- 不動産市場などで広く用いられる.
- 数式: $P = \theta_0 + \sum_{i=1}^k \theta_i X_i + \epsilon$

問題点

- **多重共線性**: 説明変数間の高い相関により回帰係数の推定が不安定になる.
- **欠落変数バイアス**: 重要な変数が欠落すると回帰係数の推定に歪みが生じる.
- **交互作用の存在**: 複数の変数が同時に影響を与える場合に交互作用項を導入することで, より精緻なモデルが可能になる.

3.1 ヘドニック法とその問題点

12/27

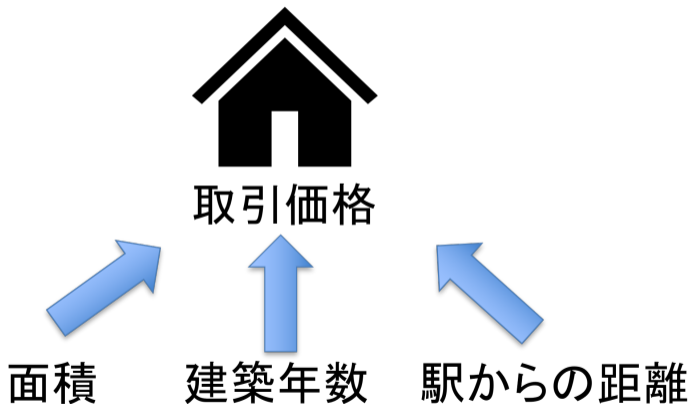
1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに



3.2 スパース推定の手法

13/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

- **Lasso 回帰**: 目的関数に L1 正則化項を加え, 不要な変数の係数を 0 にする手法. 変数選択が自動的に行われ, モデルの解釈性が向上する.
- **Ridge 回帰**: 目的関数に L2 正則化項を加える手法で, 係数を小さく抑制するが, 0 にはしない. 多重共線性を緩和し, 過学習を防ぐ.
- **Elastic Net**: Lasso と Ridge の正則化項を組み合わせた手法. 両方の効果を取り入れ, 変数選択と多重共線性の問題に同時に対処できる.
- Elastic Net は, 相関の強い変数間でパラメータの差を小さくし, 変数選択とパラメータ推定を安定化させる.
- スパース推定は, 変数選択と係数推定が一致し, 漸近的に正規分布に従うことを保証するオラクル性を持つ.

3.2 スパース推定の手法

14/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

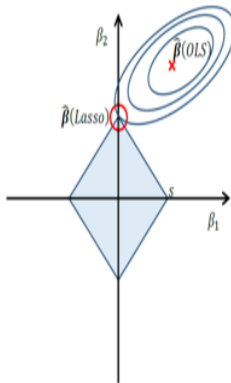
6. おわりに

Lasso

$$\operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n (Y_i - \beta_1 X_{1,i} - \beta_2 X_{2,i})^2$$

$$\text{s.t. } |\beta_1| + |\beta_2| \leq s$$

$$s > 0: \lambda \text{ と 1 対 1 対応}$$

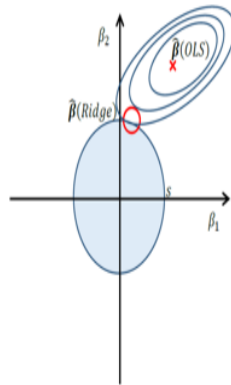


リッジ回帰

$$\operatorname{argmin}_{\beta_1, \beta_2} \sum_{i=1}^n (Y_i - \beta_1 X_{1,i} - \beta_2 X_{2,i})^2$$

$$\text{s.t. } \beta_1^2 + \beta_2^2 \leq s^2$$

$$s > 0: \lambda \text{ と 1 対 1 対応}$$



3.3 Folium を用いた Web-GIS の開発

15/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

Web-GIS とは、html 形式で記述され、World Wide Web 上でアプリケーションとして動作する GIS の総称。また、Web-GIS における html は直接手書きで作成するほかプログラミング言語で自動的に生成も可能。Web-GIS を記述する Python のモジュールを Folium と呼ぶ。

3.3 Folium を用いた Web-GIS の開発

16/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに



4.1 データセットの作成

17/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

本研究では、富山市の不動産取引データを用いて、不動産価格に影響を与える要因を分析するためのデータセットを作成する.

- e-Stat の統計データをグリッドセル単位に変換
- NAVITIME から施設の最短距離と立地数を取得
- Mapbox API を利用し土地利用の比率を計算
- 道路ネットワークを抽出し密度や平均次数を算出

これらのデータを統合し、不動産価格形成要因の分析を行う.

4.1 データセットの作成

18/27

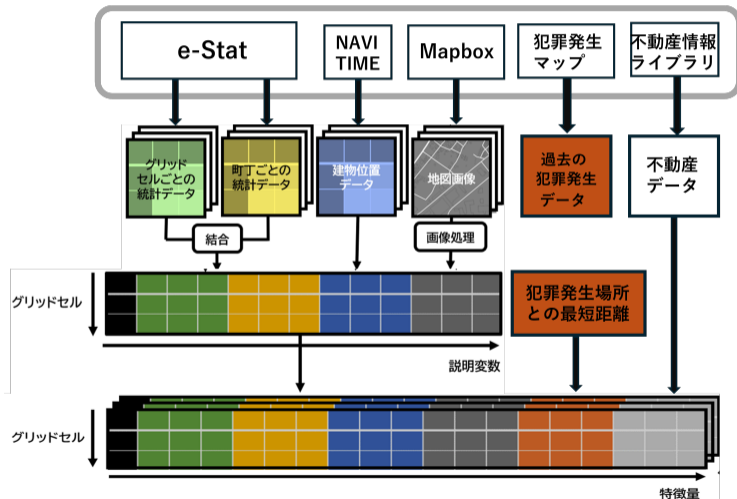
1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに



4.2 AEN による変数選択

19/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

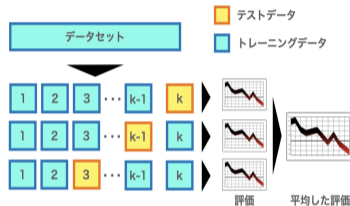
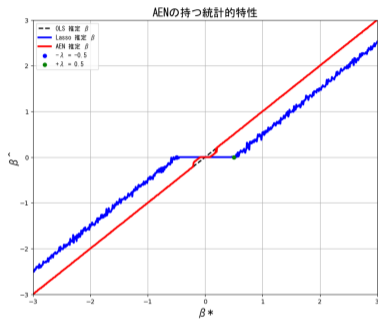
5. 数値実験並びに考察

6. おわりに

- AEN (Adaptive Elastic Net) は, LASSO と Ridge 回帰を組み合わせた手法.
- L1 正則化により変数選択を行い, L2 正則化で多重共線性を抑制.
- ハイパーパラメータ α (L1 と L2 のバランス) と λ (正則化強度) が重要.
- 交差検証により最適な α と λ を選定.

4.2 AEN による変数選択

20/27



k分割交差検証の例

$$\hat{\beta}^{1st} = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \arg \min_{\beta} \left(\|Y - X\beta\|^2 + \lambda_2 \sum_{k \geq j \geq 0} \beta_{jk}^2 + \lambda_1 \sum_{k \geq j \geq 0} |\beta_{jk}| \right) \right\}$$

$$\hat{w}_{jk} = \left(|\hat{\beta}_{jk}^{1st}| \right)^{-\gamma}$$

$$\hat{\beta} = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \arg \min_{\beta} \left(\|Y - X\beta\|^2 + \lambda_2 \sum_{k \geq j \geq 0} \beta_{jk}^2 + \lambda_1^* \sum_{k \geq j \geq 0} \hat{w}_{jk} |\beta_{jk}| \right) \right\}$$

AENの数式

$$\lambda_1 = \alpha_1 \cdot l1_ratio_1$$

$$\lambda_2 = \alpha_1 \cdot (1 - l1_ratio_1)$$

$$\lambda_1^* = \alpha_2 \cdot l1_ratio_2$$

ハイパーパラメータ

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

4.3 不動産価格形成要因の分析と GIS による可視化

21/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

ヘドニック・アプローチに基づき，一次項，二次項，交互作用項を含む回帰分析を行い，回帰係数の有意性を検定した後，得られた結果を Web-GIS で可視化し，地域ごとの不動産価格形成要因を視覚的に表現する．

5.1 数値実験の概要（1）

23/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

- 動画で数値実験の流れをお見せします.

5.1 数値実験の概要 (2)

24/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

最寄り駅名_富山	前面道路：種類_国道	最終学歴が中学以下の人口割合	駅（最短距離）
最寄駅：距離	前面道路：種類_市道	最終学歴が高校の人口割合	小学校（最短距離）
面積	前面道路：種類_私道	最終学歴が大学以上の人口割合	中学校（最短距離）
土地の形状_不整形	前面道路：種類_県道	居住年数5年未満の人口割合	高校（最短距離）
間口	都市計画_市街化調整区域	居住年数が20年以上の人口割合	大学（最短距離）
延床面積	建ぺい率	一戸建て割合	旅館（最短距離）
建築年数	容積率	アパート・低中層マンション割合	ホテル（最短距離）
建物の構造_木造	犯罪発生率	ショッピングモール（最短距離）	道路の面積比率
用途_住宅	人口	警察署／交番（最短距離）	建物の面積比率
前面道路：方位_北	18歳未満人口割合	駐車場（最短距離）	水の面積比率
前面道路：方位_北東	65歳以上人口割合	駐輪場（最短距離）	空き地の面積比率
前面道路：方位_南	外国人人口割合	金融機関（最短距離）	ノード数（道路）
前面道路：方位_南東	世帯数	保育園／幼稚園（最短距離）	エッジ数（道路）
前面道路：方位_南西	単身世帯割合	レジャー施設（最短距離）	密度（道路）
前面道路：方位_東	核家族世帯割合	スーパー（最短距離）	平均次数（道路）
前面道路：方位_西	正規労働者割合	コンビニ（最短距離）	平均次数（道路）
前面道路：幅員	非正規労働者割合	デパート（最短距離）	min_distance_to_crime

5.2 実験結果と考察 (1)

25/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

選ばれた説明変数	係数	選ばれた説明変数	係数
面積 (㎡)	-0.3931551	最寄り駅名_富山×建築年数	-0.08785
建築年数	-1.2146915	最寄り駅名_富山×18歳未満人口割合	0.00165
デパート	-0.14261367	最寄り駅名_富山×ホテル	-0.00064
面積 (㎡) ^2	0.15009339	最寄り駅名_富山×スーパー	0.02225
建築年数^2	0.62157349	最寄り駅名_富山×建物の面積比率	0.07147
面積 (㎡) × 建築年数	0.1119879	土地の形状_不整形×平均次数 (道路)	-0.00662
面積 (㎡) × 道路の面積比率	-0.02498012	建物の構造_木造×単身世帯割合	-0.03665
延床面積 (㎡) × 容積率 (%)	0.08866882	用途_住宅×面積 (㎡)	-0.00451
延床面積 (㎡) × 世帯数	0.00142986	用途_住宅×エッジ数 (道路)	0.0303
延床面積 (㎡) × 小学校	0.00834187	前面道路：方位_北×単身世帯割合	0.00781
延床面積 (㎡) × ノード数 (道路)	0.05614749	前面道路：方位_南×世帯数	0.0102
建築年数×容積率 (%)	-0.00777942	前面道路：方位_南×アパート・低中層マンション割合	0.00652
建築年数×犯罪発生率(件/25000)	-0.01601939	前面道路：方位_南×密度 (道路)	0.00237
建築年数×居住年数5年未満の人口割合	-0.07917069	前面道路：方位_南×min_distance_to_crime	6.66E-05
18歳未満人口割合×金融機関	0.0070231	前面道路：種類_県道×アパート・低中層マンション割合	0.01903
18歳未満人口割合×水の面積比率	0.0046183	前面道路：種類_私道×延床面積 (㎡)	0.03177
世帯数×居住年数5年未満の人口割合	0.02545681	前面道路：種類_国道×アパート・低中層マンション割合	0.01456
世帯数×デパート	-0.02471389	都市計画_市街化調整区域×容積率 (%)	-0.04275
単身世帯割合×ショッピングモール	0.01539941	都市計画_市街化調整区域×居住年数が20年以上の人口割合	-0.00482
正規労働者割合×高校	-0.03546675	最寄り駅名_富山×用途_住宅	0.03413
最終学歴が大学以上の人口割合×居住年数5年未満の人口割合	0.1201493	最寄り駅名_富山×前面道路：方位_南東	0.0013
居住年数5年未満の人口割合×高校	-0.00851653	最寄り駅名_富山×前面道路：種類_市道	0.04655
一戸建て割合×警察署/交番	-0.00485225	建物の構造_木造×前面道路：種類_私道	-0.02486
アパート・低中層マンション割合×min_distance_to_crime	0.01265782	前面道路：方位_南東×前面道路：種類_市道	0.00353
ホテル×道路の面積比率	-0.0215068		

・AENにより2329個の説明変数から49個に絞り込み、面積、建築年数、延床面積などの基本的な不動産要素に加え、道路の面積比率、犯罪発生率、学校が存在、最寄り駅、道路の方位など、地域や社会的要因も選ばれた。

・最寄り駅名富山や前面道路の方位、都市計画などの地域特性が価格決定に重要であることが示され、地域特性と物理的要因の交互作用が不動産価格に強い影響を与えることが確認された。

AENによって選択された変数

5.2 実験結果と考察 (2)

26/27

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに

変数	回帰係数	有意性	変数	回帰係数	有意性
定数項	10.90494217	1% 有意	ホテル×道路の面積比率	-0.02442423	5% 有意
面積 (㎡)	-0.45057681	1% 有意	最寄り駅名_富山×建築年数	0.056851774	1% 有意
建築年数	-0.75273883	1% 有意	最寄り駅名_富山×18歳未満人口割合	0.026772377	有意でない
デパート	-0.19616711	1% 有意	最寄り駅名_富山×ホテル	-0.01875637	有意でない
面積 (㎡) ^2	0.136834162	1% 有意	最寄り駅名_富山×スーパー	0.004121636	有意でない
建築年数^2	0.157313698	1% 有意	最寄り駅名_富山×建物の面積比率	-0.04567861	5% 有意
面積 (㎡) × 建築年数	0.027000603	有意でない	土地の形状_不整形×平均次数 (道路)	-0.03401206	1% 有意
面積 (㎡) × 道路の面積比率	0.062998972	5% 有意	建物の構造_木造×単身世帯割合	-0.02987425	5% 有意
延床面積 (㎡) × 容積率 (%)	0.058310345	1% 有意	用途_住宅×面積 (㎡)	0.006790478	有意でない
延床面積 (㎡) × 世帯数	-0.00751221	有意でない	用途_住宅×エッジ数 (道路)	0.045643863	1% 有意
延床面積 (㎡) × 小学校	0.049956034	1% 有意	前面道路: 方位_北×単身世帯割合	0.013193252	有意でない
延床面積 (㎡) × ノード数 (道路)	0.083565726	1% 有意	前面道路: 方位_南×世帯数	0.015804836	有意でない
建築年数×容積率 (%)	-0.0265805	有意でない	前面道路: 方位_南×アパート・低中層マンション割合	-0.00424951	有意でない
建築年数×犯罪発生率(件/25000)	-0.03107986	1% 有意	前面道路: 方位_南×密度 (道路)	0.006152642	有意でない
建築年数×居住年数5年未満の人口割合	0.052909542	1% 有意	前面道路: 方位_南×min_distance_to_crime	0.0078497	有意でない
18歳未満人口割合×金融機関	0.016574485	10% 有意	前面道路: 種類_県道×アパート・低中層マンション割合	0.015788703	10% 有意
18歳未満人口割合×水の面積比率	0.007888894	有意でない	前面道路: 種類_私道×延床面積 (㎡)	0.037926511	1% 有意
世帯数×居住年数5年未満の人口割合	0.003683871	有意でない	前面道路: 種類_国道×アパート・低中層マンション割合	0.021698973	1% 有意
世帯数×デパート	-0.01766994	有意でない	都市計画_市街化調整区域×容積率 (%)	-0.03586861	1% 有意
単身世帯割合×ショッピングモール	0.039031645	1% 有意	都市計画_市街化調整区域×居住年数が20年以上の人口割合	-0.06980964	1% 有意
正規労働者割合×高校	0.00523449	有意でない	最寄り駅名_富山×用途_住宅	-0.00825632	有意でない
最終学歴が大学以上の人口割合×居住年数5年未満の人口割合	0.055265477	1% 有意	最寄り駅名_富山×前面道路: 方位_南東	0.011028724	有意でない
居住年数5年未満の人口割合×高校	-0.07914903	1% 有意	最寄り駅名_富山×前面道路: 種類_市道	0.028617695	有意でない
一戸建て割合×警察署/交番	0.000722292	有意でない	建物の構造_木造×前面道路: 種類_私道	-0.05506079	1% 有意
アパート・低中層マンション割合×min_distance_to_crime	0.047344872	1% 有意	前面道路: 方位_南東×前面道路: 種類_市道	0.017832109	5% 有意
			決定係数 (R^2)	0.7189	

・決定係数は0.7189で、モデルは不動産価格の約71.89%を説明しており、適合度が高い。

・面積(㎡)や建築年数は有意な変数であり、面積の増加は不動産価格を減少させ、建築年数の増加も価格を減少させることが示された。また、デパートとの距離が近いほど不動産価格に負の影響を与える可能性が示唆された。

・交互作用項では、面積(㎡)と道路の面積比率の交互作用が有意であり、道路面積比率が高いほど面積の影響が強くなることが示された。また、最寄り駅名富山×建築年数や建物の構造_木造×単身世帯割合が有意であり、地域特性や建物構造が不動産価格に与える影響を反映していることが示された。

6. おわりに

27/27

まとめ

- 多数の説明変数からAENを用いて説明変数の選択を行うことができた.
- 交互作用項を考慮することで, 立地や周辺環境による価格への影響の違いを明らかにした.

今後の課題

- 説明変数の候補数が多数であったため, AENによる推定に要する時間が課題であり, 実行時間がアルゴリズムの改善や並列計算の導入が必要.
- 異なる地域や時系列データを用いた分析を行い, 本手法の一般化可能性を評価する.

1. はじめに

3. ヘドニック・アプローチによる不動産価格形成要因の分析

4. 提案手法

5. 数値実験並びに考察

6. おわりに