

## 論文

# 線形回帰モデルとニューラルネットワークによる 不動産価格推定に関する考察

植杉 大 ・ 増尾 歩

## A Study of Real Estate Price Estimation with Linear Regression Model and Neural Network Model

Dai Uesugi and Ayumu Masuo

### 【要 旨】

本論文は、線形回帰モデル、並びに 1 層ニューラルネットワークモデル及び 3 層ニューラルネットワークモデルによる不動産価格推定に関して、それぞれの特性について明らかにした。決定係数によるモデルの説明力を比較すると、3 層ニューラルネットワークモデル、1 層ニューラルネットワークモデル、単純な線形回帰モデルの順で説明力が高いという結果となったが、その差は僅差である。より複雑かつ高度な線形回帰モデルの構築により、その差はほぼなくなるものと考えられる。また、層を厚くすること(よりディープにすること)で予測精度が向上することが確認された。一般に、ニューラルネットワークモデルは予測精度の向上を目的としているため、モデルから得られる経済学的インプリケーションを引き出すのが困難であるといった見方に対して、各説明変数(インプット)の限界効果を求め、線形回帰モデルによる推定パラメータとの整合性比較の方法を提示した。

## 1. はじめに

本論文は、不動産価格推定方法として、線形回帰モデルによる推定、及びニューラルネットワークモデルによる推定に関するそれぞれの特性を明らかにすることを目的としている。

不動産価格推定を行う場合、不動産価格はその属性の関数であると考えるヘドニックモデルに基づいて、線形回帰モデルによる推定が主流とされてきた<sup>[8]</sup>。この場合の属性とは、具体的には CBD からの時間的・物理的距離、最寄り駅からの時間的・物理的距離、土地面積、延床面積、築年数等の数値データのみならず、所在する用途地域、使用目的、建物の構造、取引事情、鉄道などの沿線、取引時点など、様々な質的データとして示されるものである。

線形回帰モデルによる不動産価格推定を行う場合、不動産価格との非線形関係が認められる数値データについては対数変換等の適切な処理を行い、質的データについてはダミー変数に変換することが必要である。また、説明変数同士の相関は独立もしくは弱いと仮定され、多重共線性の問題がクリアになっていることも推定の前提条件である。さらに、単純な線形回帰モデルの誤差項は、正規分布、分散均一、系列相関なしと仮定できることなども推定の前提条件である。したがって、不動産価格推定に線形回帰モデルを用いる場合、厳密な変数処理やかなり厳しい仮定を設けなければならない。

1980 年代後半以降、不動産の空間的性質を明示的にモデルに組み込んだ SAR や SEM に代表される空間計量モデル、属性に係る推定パラメータに対して所在位置に応じて重みを与える地理的加重回帰など、空間的自己相関や空間的異質性を考慮したより洗練されたモデルが提案され、価格に対する説明力・予測能力は向上した。

以上のように、基本的な線形回帰モデルであれ、空間計量経済モデルであれ、知見に基づく仮説をモデル化し、組み込むべき変数を精査し、適切な変数変換を施し、推定パラメータが統計的に有意であるかを確認し、不動産価格に対する説明力・予測能力が十分であるかを確認するという一連の作業を通じて、仮説に関する全体的な理論的整合性を検証することを目的としている。

一方で、2000 年以降のインターネットの発展を受けたオープンデータやビッグデータの利活用、近年の不動産テックの進展を背景に、ニューラルネットワークによる不動産価格推定が行われるようになってきている<sup>[1][2]</sup>。最近ではディープニューラルネットワークによる不動産価格推定に関する研究が行われているが、実は両者の推定精度を比較するといった研究は 1990 年代から現在に至るまで多く行われており、その歴史は意外に長い<sup>[10]</sup>。

推定に用いられるデータの数、データの種類の特性、説明変数と被説明変数の非線形性などによって両者の推定精度の優劣は評価が分かれているが、概ねニューラルネットワークモデルの方が推定精度が高いといった結果が得られている<sup>[3][4][5][6][7][9]</sup>。予測確度を上げるためにモデルパラメータの学習を行うのがニューラルネットワークモデルの目的であるので、当然といえば当然の結果である。

このように、計量経済分析における線形回帰モデルとニューラルネットワークモデルはその

目的が異なっているため、推定精度だけでどちらが優れたモデルかという単純な比較をしても意味はない。むしろ、計量経済分析の中心的存在である線形回帰モデルを基準として、ニューラルネットワークモデルの長所、短所を比較、考察することに意味があると考える。そこで本論文では、線形回帰モデル、1層ニューラルネットワークモデルおよび3層ニューラルネットワークモデルの推定結果を通じて、それぞれの予測精度、モデル構築及びモデル評価の難易度、説明変数が被説明変数に与える限界効果について、比較、考察する。特に限界効果に関する分析については、われわれが知る限りこれまでに十分研究が行われていない点であり、本論文の貢献部分といえる。

本論文の構成は以下のとおりである。第2節は、分析に使用するデータについて説明する。第3節は、ベースとなる線形回帰モデル、1層ニューラルネットワークモデル、3層ニューラルネットワークモデルの推定結果を示す。第4節は、前節の結果を比較、考察する。第5節では、まとめと今後の課題を示す。

## 2. 使用データ

本論文で使用するデータは、国土交通省不動産・建設経済局が提供する土地総合情報システムから取得した、大阪府大阪市の2019年第1四半期から2020年第1四半期までの不動産取引データである。CSVファイル形式でダウンロードした取引データ項目のうち、被説明変数である取引価格のほか、最寄り駅からの時間的距離、土地面積、延床面積、前面道路の幅員、築年数等の数値データ、所在する用途地域、用途、建物の構造、取引事情といった質的データをダミー変数化したものを説明変数とした。

地域及び期間の範囲条件、項目を絞り込み、欠損値のあるデータを削除した結果、データ数は2254件となった。

数値データの中には例外的な記法でテキスト入力されているデータが存在している。これらを適切に置換し数値データに変換した。特に数値データは、被説明変数・説明変数間の非線形関係を考慮し、線形回帰分析に対応して対数変換を施した。

質的データのダミー変数化は以下のように行った。用途地域ダミーについては住宅地域をベースとして商業地域ダミー、工業地域ダミーを作成した。用途については当該不動産が現状利用を示しており様々な性質が複数入力されており、主な性質として住宅・事務所・工場・店舗を選択し、これらが含まれているものにフラグを立てた(したがって1つのデータに複数のフラグが立つ場合もある)。建物の構造についても性質が複数入力されており、主な性質として木造・RC・SRC・鉄骨造を選択し、これらが含まれているものにフラグを立てた。取引事情についても性質が複数入力されており、主な性質として関係者間取引・調停／競売・隣地の購入を選択してこれらが含まれているものにフラグを立て、特段の事情がない取引をベースとした。

以上、説明変数の作成方法については図表1にまとめた。

図表1 説明変数リスト

変数名	データ項目名	作成方法	予想される 符号条件
dist	最寄駅：距離 (分)	「30分?60分」「1H?1H30」「2H?」という入力 は、それぞれ45、75、120として変換した。	(-)
log_area	面積 (㎡)	数値を対数化した。なお、「2000㎡以上」という入力は 2000として変換した。	(+)
log_floor	延床面積 (㎡)	数値を対数化した。なお、「2000㎡以上」という入力は 2000として変換した。	(+)
road	前面道路：幅員 (m)		(+)
age	築年数(建築年)	元データ『建築年』の和暦を西暦に変換後、経過年数を 計算した。ただし「戦前」という入力は1941年として 計算した。	(-)
wood,rc, src,sf	建物の構造	データ項目に「木造」・「RC」・「SRC」・「鉄骨造」を含 む場合に1、そうでない場合には0とした。	
house,office, factory,shop	用途	データ項目に「住宅」・「事務所」・「工場」・「店舗」を含 む場合には1とした。	
commercial	地域	データ項目が「商業地」である場合に1とした。	(+)
industrial		データ項目が「工業地」である場合に1とした。	(-)
relatives	取引の事情等	データ項目に「関係者間取引」を含む場合に1とした。	(-)
arbitration		データ項目に「調停・競売等」を含む場合に1とした。	(-)
neighbor		データ項目に「隣地の購入」を含む場合に1とした。	(+)

線形回帰モデルとニューラルネットワークによる  
不動産価格推定に関する考察

被説明変数及び説明変数の記述統計量は図表 2 にまとめた。

図表 2 変数の記述統計

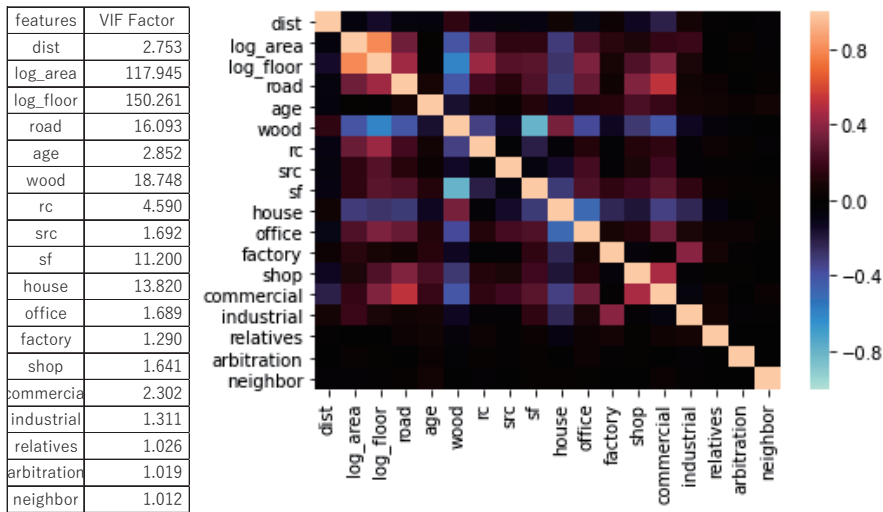
	mean	std	min	25%	50%	75%	max
log_price	17.389	1.223	13.122	16.706	17.342	17.910	21.679
dist	7.932	6.348	0.000	4.000	7.000	10.000	120.000
log_area	4.603	0.772	3.401	4.094	4.382	5.011	7.601
log_floor	5.036	0.892	2.303	4.500	4.700	5.470	7.601
road	1.834	0.604	0.000	1.386	1.792	2.079	4.382
age	26.376	21.517	0.000	1.000	26.000	44.000	79.000
wood	0.606	0.489	0	0	1	1	1
rc	0.101	0.302	0	0	0	0	1
src	0.012	0.111	0	0	0	0	1
sf	0.297	0.457	0	0	0	1	1
house	0.896	0.305	0	1	1	1	1
office	0.095	0.294	0	0	0	0	1
factory	0.019	0.137	0	0	0	0	1
shop	0.131	0.338	0	0	0	0	1
commercial	0.178	0.383	0	0	0	0	1
industrial	0.022	0.147	0	0	0	0	1
relatives	0.005	0.070	0	0	0	0	1
arbitration	0.011	0.103	0	0	0	0	1
neighbor	0.003	0.056	0	0	0	0	1

次に説明変数間の相関関係について、VIF 値及びヒートマップにより検討する(図表 3 を参照)。

線形回帰モデルでは、説明変数間の相関が高く両者の間に強い線形関係が認められる場合、多重共線性が発生する。これによりパラメータ推定が不安定になり、深刻な場合には逆行列計算が不能となりパラメータ推定が不能となる。VIF 値は一般に 10 を基準としてそれを超えると多重共線性が示唆され、線形関係が認められる変数のいずれかを除去する等の対処が行われ推定モデルの変更が求められる。

しかしこれはあくまで理想的な変数選択であり、不動産価格推定に用いられる説明変数は一般に相関関係の認められそうなものが多い。本研究でも、例えば土地面積と延床面積の関係及び木造と住宅利用(正の相関)、木造と鉄骨造(負の相関)など、明らかに相関がありそうな変数が(意図的に)選択されている。VIF 値やヒートマップを確認すると相関が強いことは明らかである。しかし多重共線性がある程度高くても逆行列計算が可能であれば推定結果は算出され、またパラメータの検定統計量である t 値も比較的高い値が求められてしまうのは経験上よくあることである。この問題を抱えてのヘドニック回帰の妥当性について、厳密な正解は今のところなく、事実上分析者の判断に任されている。

図表3 説明変数のVIF値とヒートマップ



### 3. モデル

本論文では、最もシンプルな線形回帰モデル、1層のニューラルネットワーク、および3層のニューラルネットワークの結果について、それぞれの2乗和誤差および決定係数を比較する。これを通じて線形回帰モデル及びニューラルネットワークによる不動産価格推定の特徴を検討する。

#### 3.1 線形回帰モデル

本論文における線形回帰モデルは、第2節で検討した説明変数に不動産価格を回帰する、シンプルな線形回帰モデルをベースとする。推定結果は図表4に示す。

図表4 線形回帰モデルの推定結果

	coef	stderr	t
const	13.0832	0.169	77.574
dist	-0.0148	0.002	-7.36
log_area	0.539	0.03	17.797
log_floor	0.4858	0.032	15.338
road	0.1337	0.026	5.104
age	-0.0196	0.001	-30.998
wood	-0.2699	0.098	-2.757
rc	0.0637	0.092	0.695
src	0.2811	0.148	1.898
sf	-0.2245	0.098	-2.3
house	-0.1349	0.052	-2.59

線形回帰モデルとニューラルネットワークによる  
不動産価格推定に関する考察

office	0.1109	0.053	2.089
factory	0.067	0.107	0.627
shop	0.0851	0.043	1.96
commercial	0.4771	0.045	10.674
industrial	-0.121	0.094	-1.294
relatives	-0.4631	0.164	-2.825
arbitration	-0.4966	0.115	-4.302
neighbor	0.4523	0.24	1.881
R2	0.81		
adj_R2	0.809		

図表 4 からわかるように、符号条件はほぼ想定通りである。また、決定係数は 0.81 とシンプルな線形回帰モデルであるにもかかわらず、非常に高い結果となった。しかし、若干の説明変数については t 値が低く、原因を特定するのは難しい。

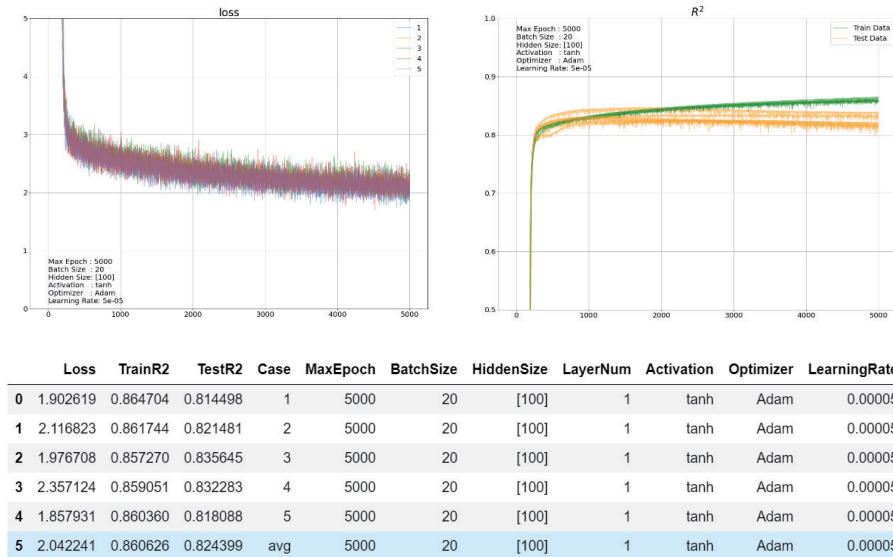
### 3.2 1 層ニューラルネットワーク

本論文では、ニューロン層の深さによって推定精度が変化するかどうかをみるために、1 層ニューラルネットワークをベースとして推定した。

はじめに、全データ 2254 件のうち 80%を訓練データ、残り 20%をテストデータとしてランダムに分割した結果、ニューラルネットワークの重みパラメータを推定するための訓練データは 1803 件、予測精度を確認するためのテストデータは 451 件となった。

本論文に示す推定結果は、ハイパーパラメータの設定として最大エポック数を 5000、バッチサイズを 20、1 層あたりのニューロン数は 200、活性化関数は tanh 関数、最適化手法は Adam、最適化の学習率は 5e-05、試行回数は 5 としたものである。推定結果は図表 5 に示す。

図表 5 1 層ニューラルネットワークの損失関数(左)と決定係数(右)の推移と試行ごとの推定結果(下)



推定結果を示した表の最下段(Case が avg の行)には、5 回の試行回数の平均が示されている。決定係数の推移のグラフと訓練データの TrainR2 をみると多少の過学習の可能性が認められるが、予測精度を見る TestR2 については各試行ともにおおむね妥当な値を示している。平均的にみると TestR2 は 0.824 となっており、線形回帰モデルと比較すると 0.015 程度の予測精度の改善がみられる。

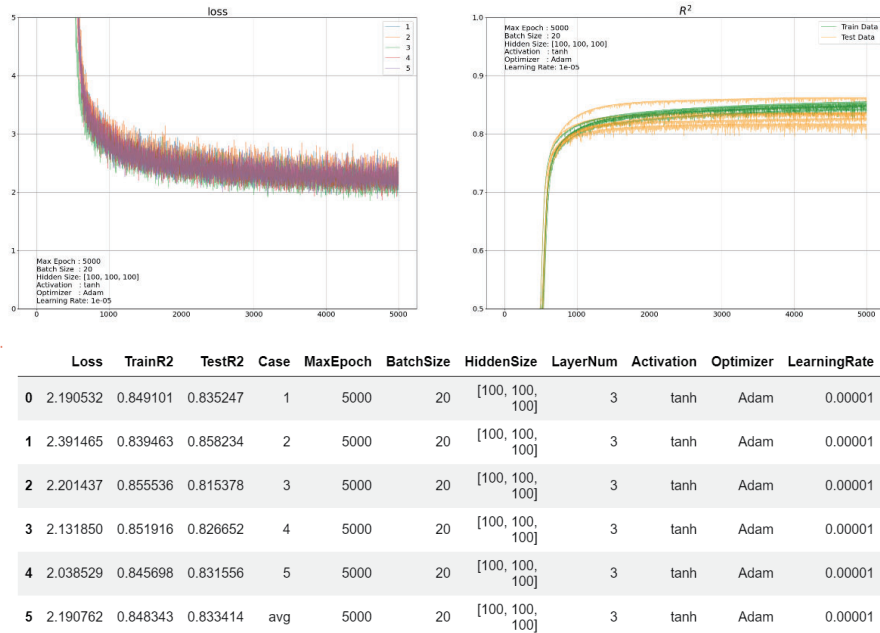
### 3.3 3 層ニューラルネットワーク

ニューロン層の深さが推定精度を向上させるかどうかをみるために、本論文では 3 層ニューラルネットワークを構築し、比較を行う。本論文に示す推定結果は、ハイパーパラメータの設定として最大エポック数を 5000、バッチサイズを 20、1 層あたりのニューロン数は 200、活性化関数は tanh 関数、最適化手法は Adam、最適化の学習率は 1e-05 としたものである。推定結果は図表 6 に示す。



線形回帰モデルとニューラルネットワークによる  
不動産価格推定に関する考察

図表 6 3層ニューラルネットワークの損失関数(左)と決定係数(右)の推移と試行ごとの推定結果(下)



1層ニューラルネットワークの場合と同様に、推定結果を示した表の最下段(Case が avg の行)には、5回の試行回数の平均が示されている。決定係数の推移のグラフと訓練データの TrainR2 をみると過学習はそれほど認められず、予測精度を見る TestR2 については各試行ともにおおむね妥当な値を示しており、特に Case2 においては訓練データよりもテストデータにおける決定係数の方が高くなっている(0.858)。平均的にみると TestR2 は 0.833 となっており、線形回帰モデルと比較すると 0.024 程度の予測精度の改善がみられる。

以上の結果から、モデルの予測精度の高さという観点で見れば、3層ニューラルネットワーク、1層ニューラルネットワーク、線形回帰モデルという順位となり、従来の研究における報告と整合的であることがわかる。

#### 4. 考察

前節での実験結果を受けて、本節では結果の検討を行い、線形回帰モデル及びニューラルネットワークモデルの特性について考察する。

##### 4.1 予測精度について

まずモデルの予測精度について考察する。本研究では予測精度について線形回帰モデルより

もニューラルネットワークモデルの方が高くなる結果となった。これは既存研究の多くで得られた結果と同じである。つまり、予測精度については総じてニューラルネットワークモデルに軍配が上がるといえるだろう。

ただしニューラルネットワークモデルによる予測精度の改善の程度については、線形回帰モデルと比較して期待されるほど大きくはない。これについても既存研究と同様の傾向がみられる。

両モデルの特性という観点からみると、線形回帰モデルと比較したニューラルネットワークモデルの魅力として、①説明変数と被説明変数間の非線形関係を捉えて予測値を算出することができる、②説明変数間で相関関係が認められる場合にそれらを含めてモデル化し、かつそのような場合に計算困難となる逆行列計算を回避することができる、という特性が挙げられる。さらに、③そのような結果を1層のニューロン層もしくは多層ニューロン層の組み合わせで実現できる実装面での手軽さ、簡便さ、拡張性の高さ、なども魅力として挙げられるだろう。

本研究の線形回帰モデルは、一部の説明変数間に(逆行列計算可能な程度の)弱い多重共線性を含んでおり、ヘドニック回帰の説明変数を選択する場合に不可避な典型的なパターンを想定している。これは理論的に線形回帰モデルの説明力を下げる一方、それと比較してニューラルネットワークモデルの予測精度の改善の程度を大きくするはずである。したがって、ニューラルネットワークモデルの予測精度の改善の程度がそれほど高くないということは、説明変数と被説明変数の非線形性が予想ほどには認められないことに原因があるのではないだろうか。

実際に、従来の不動産価格推定を目的とした線形回帰モデルにおいて変数の対数化を行い、説明変数と被説明変数の非線形関係の緩和しモデルフィットを図る工夫は必ずといっていいほど行われている。本研究でも、説明変数の一部に対してそのようなトリートメントを施しており、結果として線形回帰モデルの決定係数は0.8程度となっていることが確認されている。もちろんこれらの対処法で十分というわけではないが、ニューラルネットワークモデルの予測精度が高いことよりむしろ、これまで蓄積された推定手続きの洗練による線形回帰モデルの予測精度の高さの方が驚きに値する。

不動産価格推定においては、単純な線形回帰モデルを超えた範囲に非線形関係や変数間の複雑な関係が潜んでおり、その一部をモデル化することにより不動産計量経済分析が発展してきたといえる。それはシンプルなOLSから始まり、不動産特有の適切な説明変数選択やデータマージング、空間特性を反映したモデルなどの発展をみながら、パラメトリックなモデルの構築及びその検定手続きが確立されてきた。本研究での線形回帰モデルは最もシンプルなOLSをベースとしたために決定係数に0.05程度の差が生じたが、より洗練された計量経済モデルと比較した場合、両者の予測精度はより縮まるものと思われる。

以上の考察から、予測精度は線形回帰モデルよりもニューラルネットワークモデルの方が一般に高い傾向があるが、モデルの非線形性を適切に前処理した場合両者にはそれほど明白な差は認められないといえる。

#### 4.2 モデル構築及びモデル評価の難易度

ニューラルネットワークモデルを構築する場合、活性化関数の選択、最適化手法の選択、重みパラメータの初期値設定、ハイパーパラメータの設定方法など、様々なオプションの選択を行う必要がある。また、ニューラル層のノード数どのように決定するのか、多層のディープニューラルネットワークの場合何層重ねればいいのかなども検討が必要となる。

さらに、訓練データが少ない場合やニューロン数を多くした表現力の高いモデルの場合に過学習が発生するため、テストデータの予測精度は低下してしまう。過学習させないようにするためには、重みパラメータの初期値設定、Weight Decay、Dropout 等のテクニックを利用する必要がある。

最終的には、様々な可能性から予測精度の最も高いオプションを探索的に決定するのであるが、それは非常に time-consuming かつスキルが必要な作業である。本研究ではビッグデータ分析といえるほど十分なデータ数は確保されていないので、ニューラル層のノード数を増やすぎると過学習が発生しやすかった。また、活性化関数の種類によっては誤差逆伝播法により求めた勾配が消失しやすくなった。さらに最適化手法によって収束の仕方が大きく異なり、場合によっては収束しない場合も発生した。このようにニューラルネットワークモデルの構築には技術的熟練が必要である。

一方線形回帰モデルの場合、モデル全体の説明力(決定係数等)や各推定パラメータの検定方法(t 検定、F 検定)など様々な面から評価が可能であり、その統計学的方法是確立されている。求められた推定パラメータの経済学的インプリケーションの解釈のしやすさも線形回帰モデルの魅力である。

以上、スクラッチからニューラルネットワークモデルを構築する場合と比較して、モデル構築過程の透明性や最終的なモデル評価の容易さについては、線形回帰モデルに分があると考えらる。

#### 4.3 説明変数が被説明変数に与える限界効果について

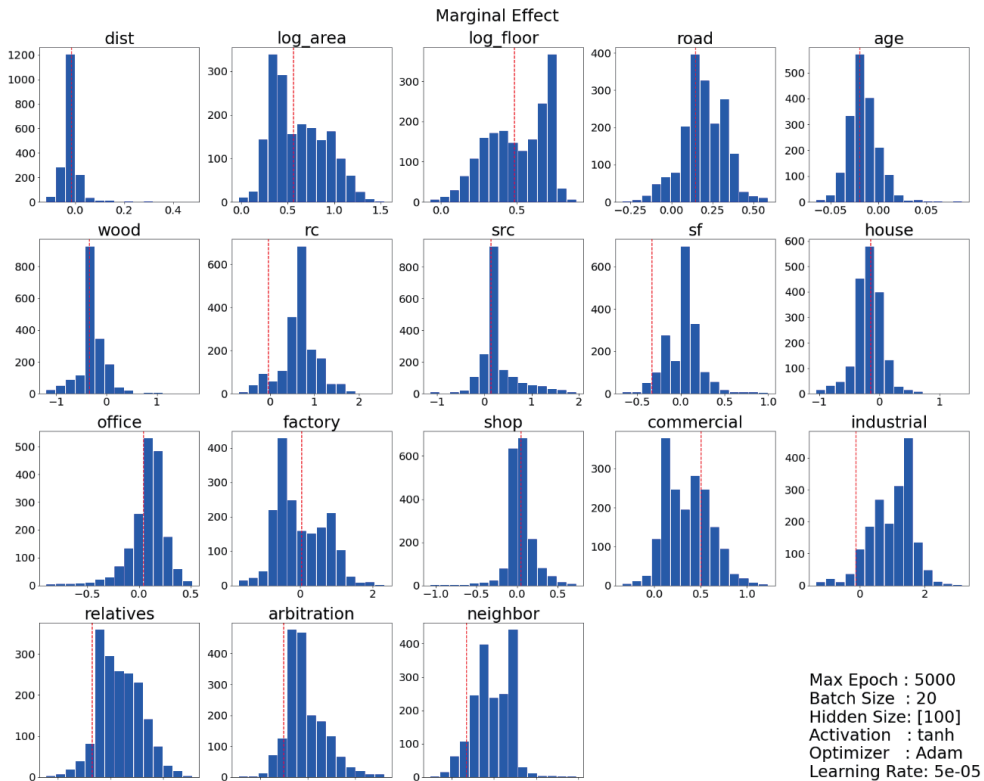
ニューラルネットワークモデルは、教師データとの予測誤差を最小にするような各ノードの重みパラメータを誤差逆伝播法及び反復的な最適化手法により求め、最終的に予測精度を可能な限り高めることを目的としている。結果として、ニューラル層のニューロン数や重ねた層の数が多ければ多いほど、重みパラメータの解釈は困難となる。一方、線形回帰モデルにおける各説明変数に係る推定パラメータは被説明変数への限界効果を表すものであり、経済学的インプリケーションを含む重要な推定量である。これらがニューラルネットワークでは求めにくいいため、ニューラルネットワークはしばしば「ブラックボックス」といわれる。

しかし、構造的に見て線形の重回帰モデルは「0 層のニューラルネットワークモデル」であって、他変数との相互作用を処理する各ノードの結合と出力がないだけである。線形回帰モデルのようにわかりやすく推定することはできないが、微分の連鎖率を用いれば、比較的簡単に各変数の限界効果が求められる。各データの他の変数との関係を含んだものとなるため、各デー

タによって限界効果が異なり、結果として分布が形成される。線形回帰モデルの推定量との比較を行うことで、他変数との複雑な関係がどの程度反映されている限界効果なのかをビジュアルで理解できる。

まずは1層ニューラルネットワークにおける各説明変数に対応する限界効果を、図表7で示す。紙面の都合上、5回の試行回数のうちTestR2が最大であった試行のみを掲載する。なお、ヒストグラムで表されているのがニューラルネットワークで求められた各地点の各説明変数に関する限界効果の頻度であり、点線で表されているのが線形回帰モデルによって求められたパラメータ推定量である。

図表7 1層ニューラルネットワークの限界効果 (Case3)

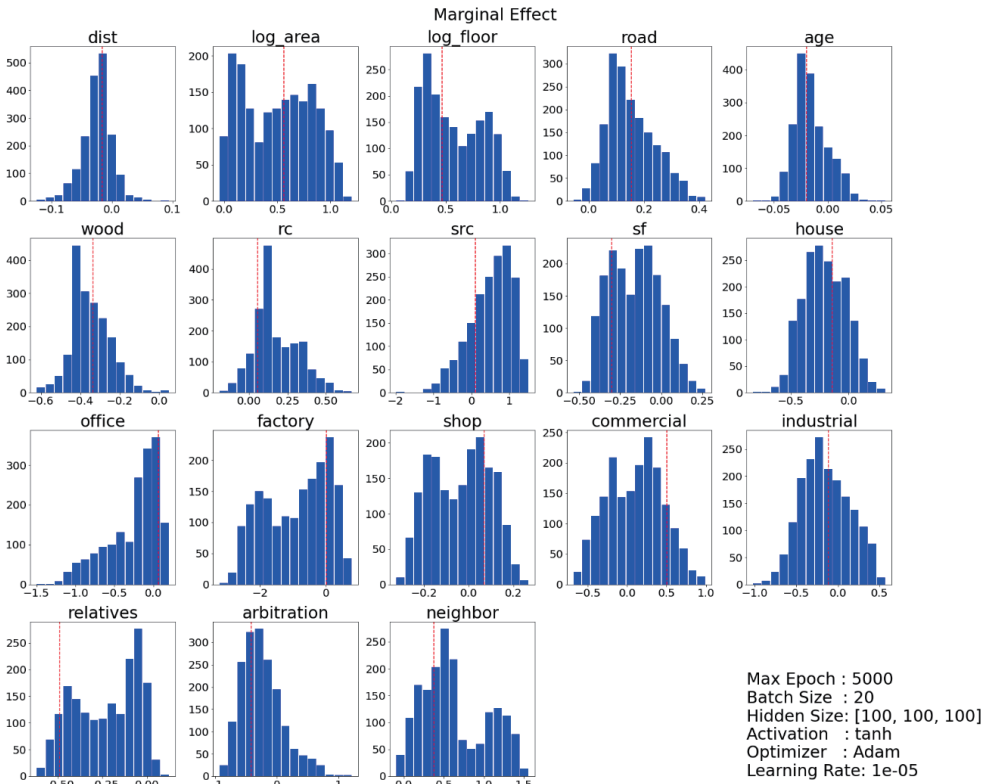


多くの説明変数において、線形回帰モデルのパラメータ推定量とニューラルネットワークの限界効果の分布のピークがほぼ一致していることがわかる。例えば線形回帰モデルで必ずといっていいほど用いられる駅からの距離(dist)や築年数(age)などは、線形回帰モデルのパラメータ検定でも  $t$  値は高く、ニューラルネットワークのヒストグラムの分布の分散は小さいことが読み取れる。しかし、ヒストグラムが左右どちらか一方に偏っていたり、多峰型になっていたりと

ると、必ずしも頻度のピークと線形回帰モデルの推定量が一致していないことが多いことがわかる。場合によっては、両者に乖離が生じている説明変数もいくつかある。

次に 3 層ニューラルネットワークの場合を見るため、結果を図表 8 に示す。この場合も紙面の都合上、5 回の試行回数のうち TestR2 が最大である Case2 の結果を示している。

図表 8 3 層ニューラルネットワークの限界効果 (Case2)



1 層ニューラルネットワークと比較してみると、限界効果の分布形状が若干異なっているものの、線形回帰モデルのパラメータ推定量との位置関係も概ね同傾向であることがわかる。

このニューラルネットワークにおける限界効果は他の変数の影響も部分的に含まれているため、線形回帰モデルの推定量とは当然異なり、データごとにも異なる値となる。しかし両者を比較すると概ね同じ領域に存在していることがわかる。むしろ分布として各データで異なる値を持っているニューラルネットワークモデルの方が、個別の何らかの要因を反映させることで高い予測精度をもたらしていると考えられる。

各説明変数が被説明変数にどの程度限界的な影響を与えるかは、重要な経済学的インプリケーションを与えるものである。したがって、線形回帰モデルの各説明変数の係数パラメータは

モデル分析で重要な役割を果たす。線形回帰モデルはこれらを明示的に求めることが目的であり、その統計的検定方法も確立されている。一方ニューラルネットワークモデルは予測精度の向上を目指すことが目的であるため、ブラックボックス内で計算される限界効果については十分な注意が払われず、経済学的インプリケーションを引き出すには不十分であるといった印象を持たれがちである。

しかし本論文で提示した方法を用いれば、他の変数を介した影響を含んでしまうものの、データごとに異なる限界効果を得ることができるため、分布として限界効果を考察することができる。その求め方も誤差逆伝播法に基づいて各レイヤに組み込んだ逆向きの計算を用いることで、プログラミング的に簡単に導出することができる。経済学的インプリケーションを引き出す手段として、このような分析方法を提示したのは本論文の貢献であるといえる。

## 5. まとめと今後の課題

本論文は、不動産価格推定方法として、線形回帰モデルによる推定、並びに1層ニューラルネットワークモデル及び3層ニューラルネットワークによる推定に関するそれぞれの特性について明らかにした。

決定係数によるモデルの説明力の観点から比較すれば、3層ニューラルネットワーク、1層ニューラルネットワーク、単純な線形回帰モデルの順で説明力が高いという結果となったが、その差は必ずしも大きくないことがわかった。より複雑かつ高度な線形回帰モデルの構築により、その差はほぼなくなるものと考えられる。また、層を厚くすること(よりディープにすること)で予測精度が多少向上することが確認された。

一方ニューラルネットワークモデルは予測精度の向上を目的としているため、モデルから得られる経済学的インプリケーションを引き出すのが困難であるといった見方に対して、各説明変数(インプット)の限界効果を求め、線形回帰モデルによる推定パラメータとの整合性比較の方法を提示した。その結果、限界効果に関してよりリッチな解釈の可能性を示唆した。

本論文では紙面の都合上、ハイパーパラメータや活性化関数、最適化手法等の組み合わせごとに総当たりで実施したニューラルネットワークモデルの結果を述べるができなかった。これらすべての実験結果から見ても、また既存研究の結果から見ても、上記結論はロバストなものといえる。特に線形回帰モデルを用いて不動産価格推定を行う場合、非線形性や多重共線性などを回避するよう、綿密・精緻に変数を作成しなければならない。したがって、同じデータを使った場合、多層ニューラルネットワークモデルの方が線形回帰モデルより予測精度が圧倒的に高くなるということは、おそらく起こりにくいのではないだろうか。今後は、同じデータを使用してどちらのモデルの精度が高いのかという比較を行うことより、使用するデータの性質等を考慮してどのモデルを用いるのかを考えるべきであろう。特にビッグデータやテキスト・画像などの非構造化データの利用を前提とするならば、データの諸性質・諸問題に対してフレキシブルに推定ができるニューラルネットワークモデルは魅力的な選択肢である。加えて、本



論文で提示した限界効果の分析方法を用いるのであれば、ニューラルネットワークモデルに対する「ブラックボックス」批判は当たらないであろう。今後は様々なデータソースを用いながら、それに適したモデルを用いて不動産価格推定の新しい方法を模索したい。

#### 参考文献

- [1] 福井光・阪井一仁・南村忠敬・三尾順一・木下明弘・高田司郎(2018), 「レインズのニューラルネットワークを用いた不動産価格査定について」, The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018.
- [2] 前田翠・関本義秀・瀬戸寿一(2019), 「異なるタイプの不動産データを用いた不動産価値推定モデルの研究〜ヘドニックと Deep Neural Network の比較」, 日本都市計画学会都市計画論文集, Vol.54, No.3, 2019 年 10 月.
- [3] Julia Margarita Núñez-Tabales, Francisco José Rey, José M Caridad (2013) 'Artificial neural networks for predicting real estate prices', Revista de Metodos Cuantitativos para la Economia y la Empresa(15), pp.30-44.
- [4] Limsombunchai, V., Gan, C. and Lee, M. (2004) 'House price prediction: hedonic price model vs. artificial neural network', American Journal of Applied Sciences, Vol. 1, No. 3, pp.193-201.
- [5] Morano, P. and Tajani, F. (2013) 'Bare ownership evaluation. Hedonic price models vs. Artificial neural network', International Journal of Business Intelligence and Data Mining, Vol. 8, No. 4, pp.340-362.
- [6] Nguyen, N. and Cripps, A. (2001) 'Predicting housing value: a comparison of multiple regression analysis and artificial neural networks', Journal of Real Estate Research, Vol. 22, No. 3, pp.313-336.
- [7] Peterson, S. and Flanagan, A.B. (2009) 'Neural network hedonic pricing models in mass real estate appraisal', Journal of Real Estate Research, Vol. 31, No. 2, pp.147-164.
- [8] Rosen, S. (1974) 'Hedonic prices and implicit markets: product differentiation in pure competition', Journal of Political Economy, Vol. 82, No. 1, pp.34-55.
- [9] Selim, H. (2009) 'Determinants of house prices in Turkey: hedonic regression versus artificial neural network', Expert Systems with Applications, Part 2, Vol. 36, No. 2, pp.2843-2852.
- [10] Worzala, E., Lenk, M. and Silva, A. (1995) 'An exploration of neural networks and its application to real estate valuation', Journal of Real Estate Research, Vol. 10, No. 2, pp.185-201.