A MODERN APPROACH TO LAND VALUATION: AN APPLICATION OF ARTIFICIAL NEURAL NETWORKS

Zhou Yang

Robert Morris University

Motivation

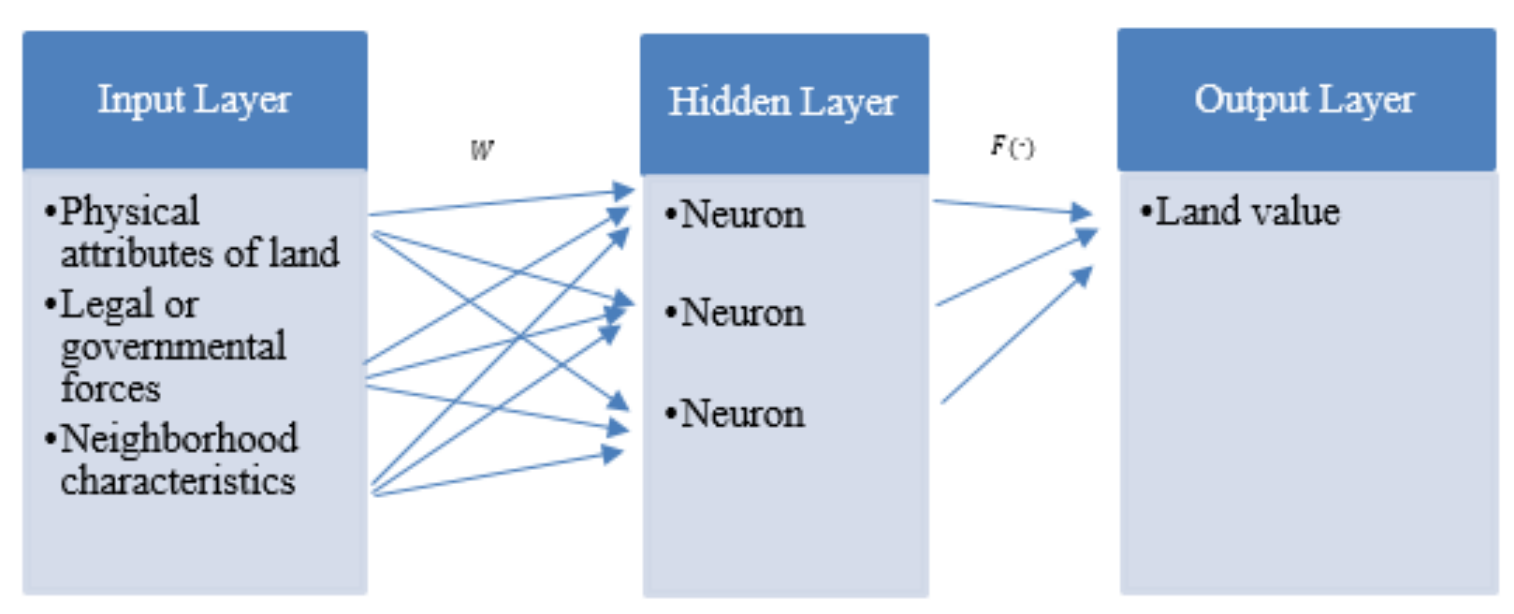
- Artificial intelligence (AI) techniques have gained increasing attention over the past few years.
- Artificial neural networks (ANN) have been used for various purposes, including prediction, pattern recognition, process control and data analysis.
- The advantage of the ANN approach is that it can examine complex nonlinear relationships without many of the parametric restrictions and a priori assumptions about the data and functional form used in conventional regression models.
  - *A growing literature on the use of ANN in real estate research (Peterson and Flanagan 2009; Lin and Mohan 2011; Zurada, Levitan and Guan 2011; McCluskey et al. 2013; Yacim and Boshoff 2018)*

Background — ANN

- Artificial neural networks provide a learning system consisting of a large number of neurons.
  - *Each neuron can process signals received and signal neurons connected to it.*
  - *With neurons organized in interconnected layers, ANNs can emulate almost any function and solve various problems, assuming sufficient training data and computing power.*
- The power of a neural network is determined by the activation functions, the learning rule, and its architecture.

Background — ANN Structure

Figure 1: An Example of ANN Architecture in Real Estate Studies



Background — ANN Training Process and Algorithms

- Available data divided into three sets: one for training, one for validation and the rest for testing.
  - *Using the training dataset to train the neural network*
  - *Utilizing the validation dataset to evaluate the model fit and find the best model fit*
  - *Reserving the test dataset for an unbiased evaluation of the selected model*
- Algorithms used for training purposes: back propagation (BP), Levenberg-Marquardt (LM), and conjugate gradient (CG) algorithms
  - *The BP algorithm: the most commonly used one in real estate mass appraisal based on ANN models (Yacim and Boshoff 2018)*

Application — Land Valuation

- Application of ANN for land value estimation
  - *Motivated by sales comparison approach*
  - *Using vacant land sales data to build ANN models, then choosing the best model to estimate land values for parcels in the improved sales dataset*
- Data preparation for ANN training
  - *Data cleaning and preprocessing (missing values; outliers)*
  - *Treatment of categorical variables (entity embeddings of categorical variables)*
- Data split for ANN training and testing
  - *No consensus about the data division ratio in the literature (avoid underfitting as well as overfitting the model)*
  - *80% used for training purposes, and 10% for validation and testing, respectively*

Application — Building ANN Models

- A multiple-layer network
  - *Multiple-layer neural networks overcoming the limitation associated with single-layer neural networks*
  - *Based on land value determinants in the existing literature and data availability, 45 variables used as inputs representing parcel characteristics as well as zoning regulations*
- Two main hyperparameters controlling the network architecture
  - *The number of layers*
  - *The number of nodes in each hidden layer*
  - *Configuring the hyperparameters through systematic experimentation with a robust test harness*

Application — Building ANN Models

Figure 2a: Predictions with One Hidden Layer

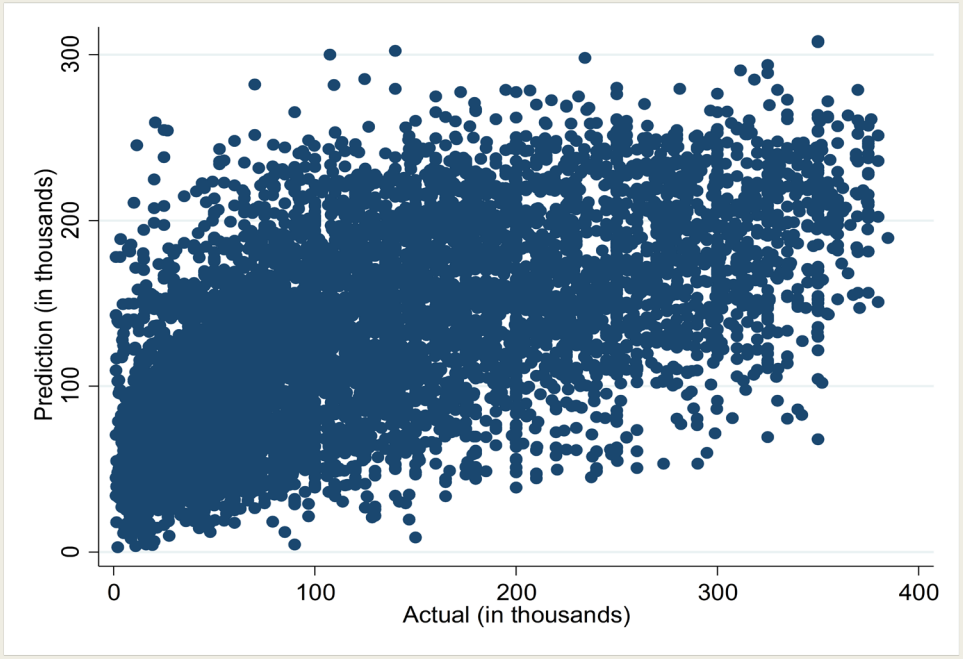
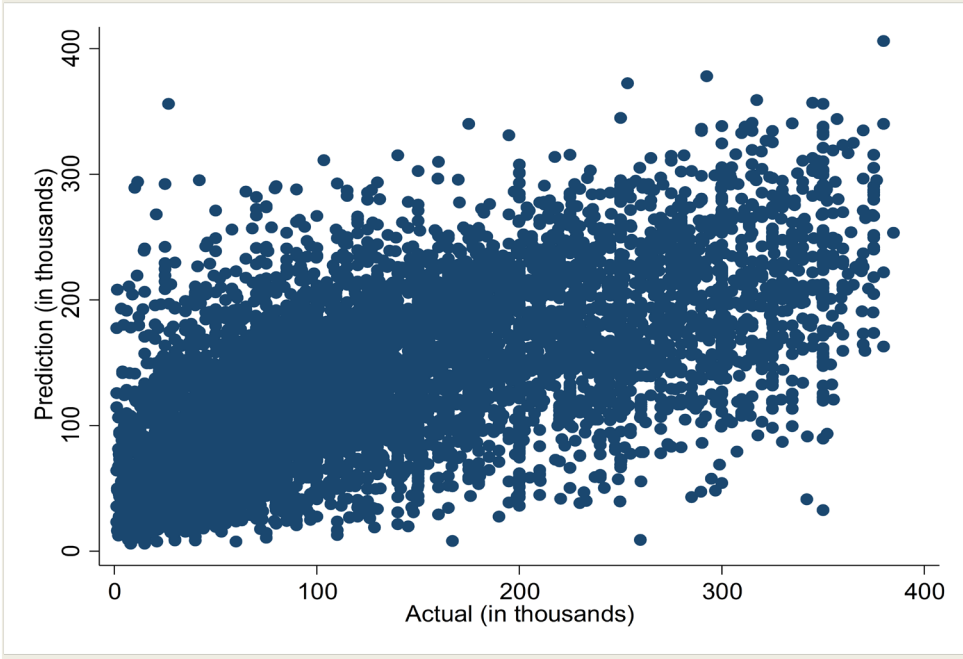


Figure 2b: Predictions with Two Hidden Layers



Application — Building ANN Models

- Activation functions
  - *Creating complex and nonlinear mappings between input and output variables*
- Popular activation functions
  - *Sigmoid activation function: it has an S-shaped curve with a range between 0 and 1 (easy to understand and apply; vanishing gradient problem and slow convergence).*
  - *Hyperbolic tangent function (Tanh): it makes output zero centered because of its range between -1 and 1 (normally preferred over Sigmoid function; vanishing gradient problem).*
  - *Rectified linear unit activation function (Relu): it rectifies the vanishing gradient problem. One potential issue is the possibility of dead neurons (leaky Relu as a fix).*

Application — Building ANN Models

Figure 3a: Predictions with Sigmoid Activation Function

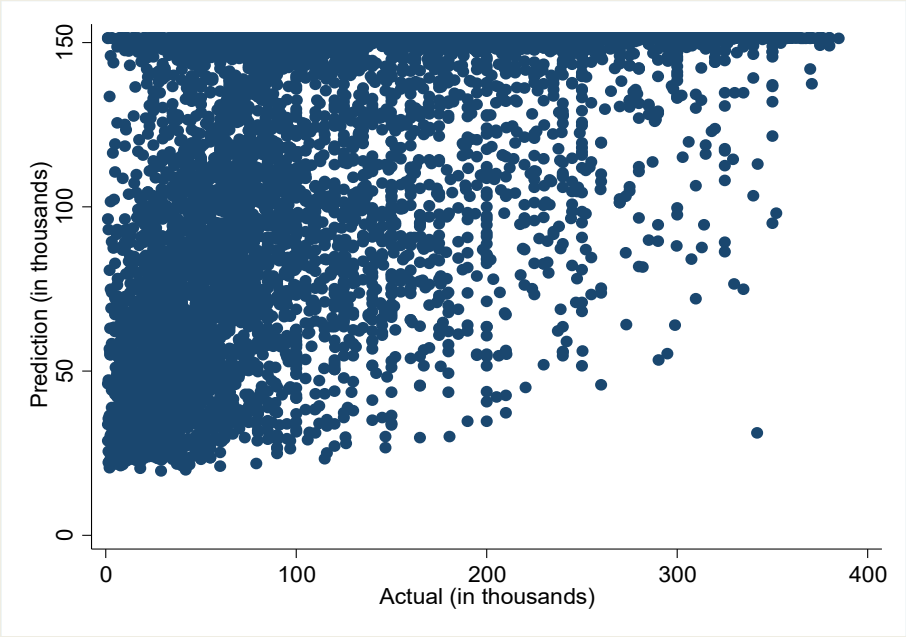
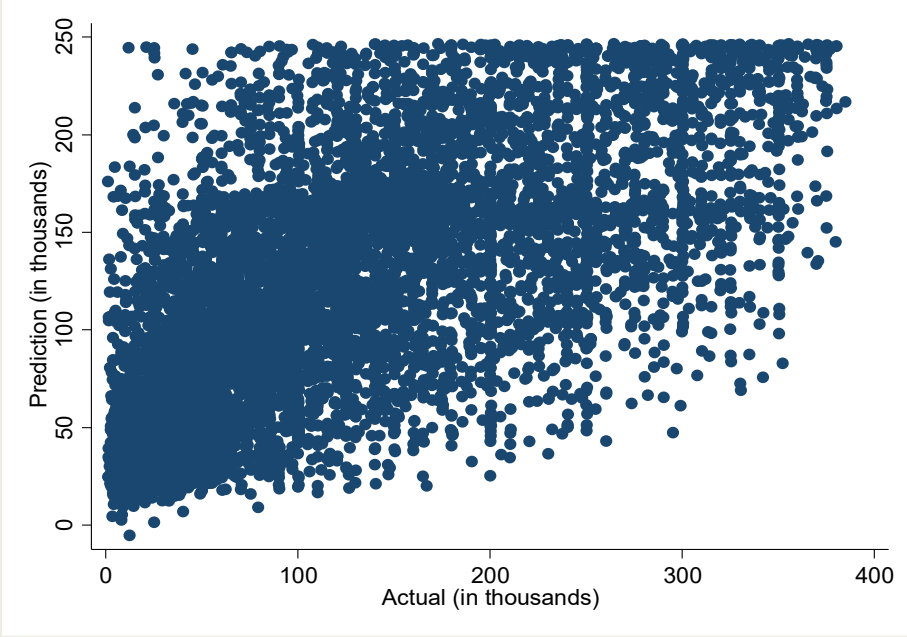
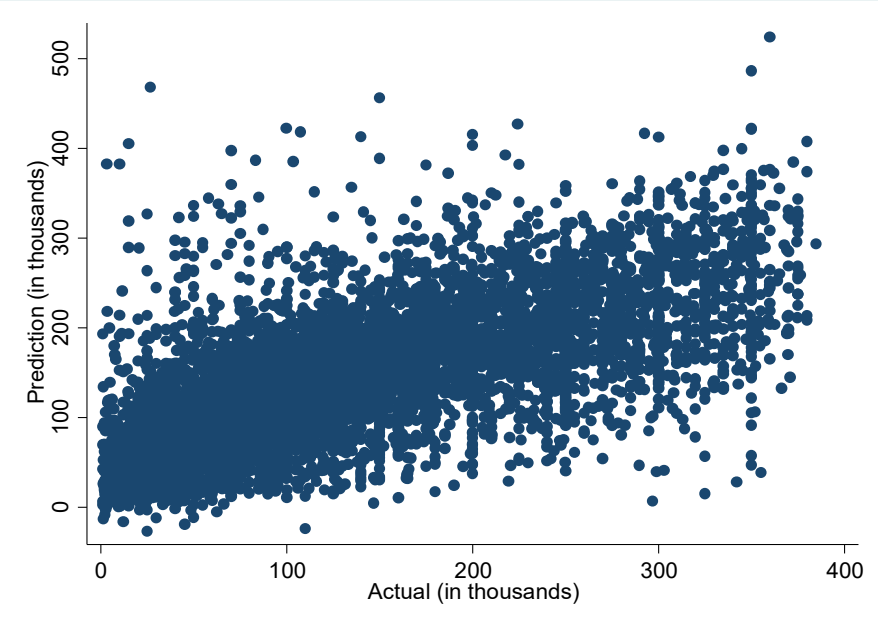
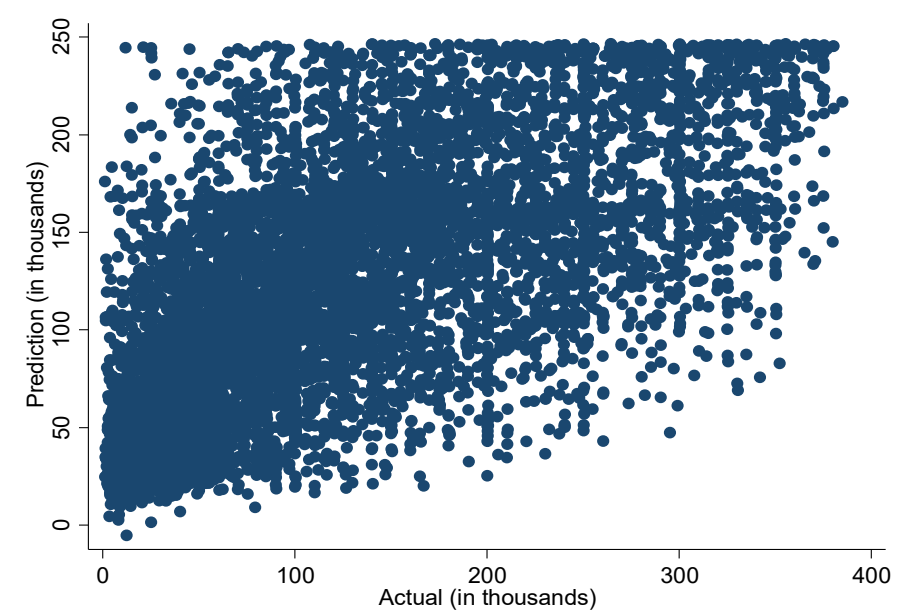


Figure 3b: Prediction Using Tanh Activation Function



Application — Building ANN Models

Figure 3c: Predictions with Relu Activation Function	Figure 3b: Predictions with Tanh Activation Function
	

Application — Building ANN Models

- In general, experimentation with ANN model configuration needs to consider efficiency as well as model capacity.
- Evaluation using the test data suggests:
  - *The prediction accuracy varies across different residential and commercial market areas.*
  - *The prediction accuracy varies across cities.*

Application — Predictions Using New Input Information

- Predict land values for parcels in the improved sales dataset
  - *Establish land values separately from property values with the constructed ANN model*
  - *Use parcel information from the improved sales dataset as new inputs to predict land values*
- Possibility to estimate land values for parcels in the county assessment dataset
  - *Similar prediction process with a few assumptions*

Limitations

- There are general concerns about out-of-sample predictions.
  - *The prediction accuracy depends on the quality of land sales data as well as the representativeness of the data.*
  - *Underrepresentation of certain land types would affect the predication accuracy for the types of land. (This problem can be mitigated as more land sales data become available over time.)*
  - *Certain land features may be important but not observable. The constructed neural networks would not be able to capture those.*

Conclusion

- ANNs provide us with a modern approach to land value estimation. One main advantage is that these models can study complex nonlinear relationships without a priori assumptions about the data and functional form.
- They can help lessen the burden on assessors or researchers when estimating land values on a large scale.
- ANNs rely on a large amount of training data to appropriately solve various problems. Inadequate information would result in poor predictions.
- Potential overfitting problems affect the generalizability of ANN models. Building a model that generalizes well to new data is always challenging.