Prediction of house prices using neural network regression models

Saatwik Bisaria

Machine learning Intern

www.ai-techsystems.com
New Delhi,India
saatwik6656@gmail.com

Abstract—**House prices increase every year, so there is a need for a system to predict house prices in the future. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.**

I. Introduction

The real estate market is one of the most competitive in terms of pricing and the same tends to vary significantly based on a lot of factors, hence it becomes one of the prime fields to apply the concepts of machine learning to optimize and predict the prices with high accuracy. Neural network regression models are being used in this project for prediction of house prices. The models being used are of a 3 layer neural network regression, 5 layer neural network regression, and a keras regressor 3 layer model.

A. Dataset

The dataset being used in this project is available on kaggle. It consists of 3000 records with 80 parameters that have the possibility of affecting the property prices. Some of the parameters are Area in square meters, Overall quality which rates the overall condition and finishing of the house, Location, Year in which house was built, Numbers of Bedrooms and bathrooms, Garage area and number of cars that can fit in garage, swimming pool area, selling year of the house and Price at which house is sold. The SalePrice is the label which we have to predict through regression techniques. Some parameters had numerical values while some had categorical values

B. Data Preprocessing

It is a process of transforming the raw, complex data into systematic understandable knowledge. It involves the process of finding out missing and redundant data in the dataset. Entire dataset is checked for NaN and whichever observation consists of NaN will be deleted. Thus, this brings uniformity in the dataset.

C. Keras

Keras is an Open Source Neural Network library written in Python that runs on top of Tensorflow. It is designed to be modular, fast and easy to use.

Keras doesn't handle low-level computation. Instead, it uses another library to do it, called the "Backend. So Keras is high-level API wrapper for the low-level API, capable of running on top of TensorFlow.

Keras High-Level API handles the way we make models, defining layers, or set up multiple input-output models. In this level, Keras also compiles our model with loss and optimizer functions, training process with fit function. Keras doesn't handle Low-Level API such as making the computational graph, making tensors or other variables because it has been handled by the "backend" engine.

MAE calculate the average of absolute error for each predicted result. MAE is useful when measuring errors in certain units. MAE values can be calculated using

$$.MAE = \sum (x_i - x)./n \qquad (2)$$

RMSE is used to calculate predicted performance by considering the prediction error of each data. RMSE formula can be seen there –

$$RMSE = \sqrt{\ } \sum (x_i - x)\text{^}2./n \qquad (3)$$

Accuracy is the measure of performance of the model. It tells how well the model is predicting the house pricing.

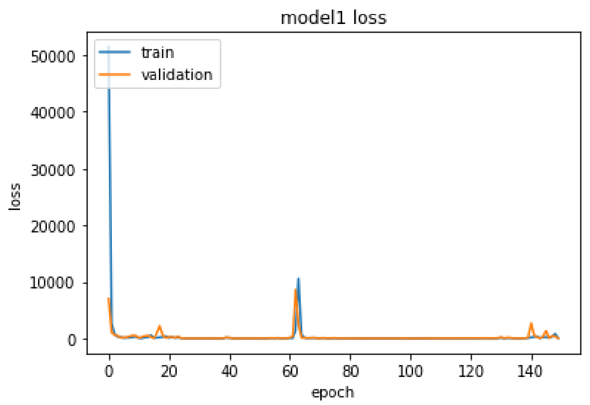
$$Accuracy=100-\sum(ypred - y)./y*100 \qquad (4)$$

II. Neural Network modelling

After preprocessing the data and converting it into usable form. 3 different models are applied on the data to identify the best model suitable for prediction of house pricing.

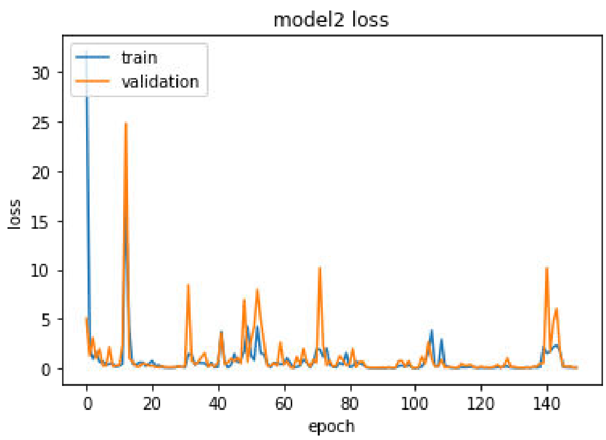
A. MODEL 1 (3 layer neural network):

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 32)	8416
dense_1 (Dense)	(None, 16)	528
dense_2 (Dense)	(None, 1)	17



B. MODEL 2 (Keras regressor based) :

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 32)	8416
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 1)	17



C. MODEL 3 (5 layer neural network):

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 128)	33664
dense_7 (Dense)	(None, 128)	16512
dense_8 (Dense)	(None, 32)	4128
dense_9 (Dense)	(None, 16)	528
dense_10 (Dense)	(None, 1)	17

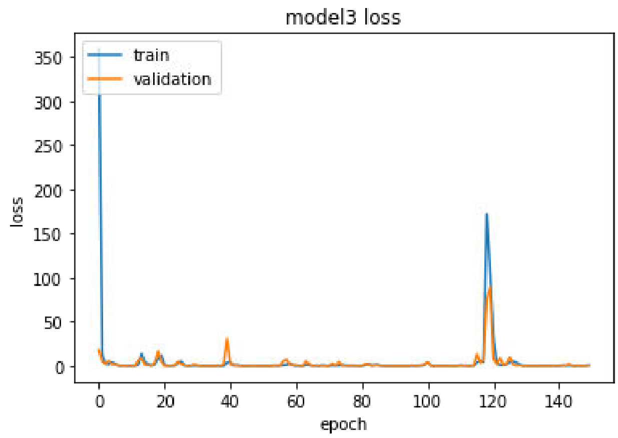


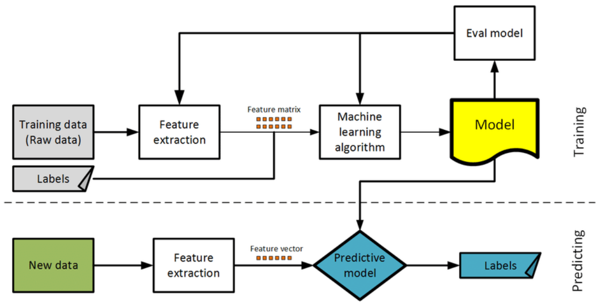
TABLE I.

MODELS	**metrics**			
	Accuracy	***Loss***	***Mean Squared Error***	***Mean absolute error***
MODEL 1 (3-Layer)	94.66815680044856	0.4521	0.4521	0.5474
MODEL 2 (3-Layer KerasRegressor)	98.83893840617066	0.0878	0.0878	0.2041
MODEL 3 (5-Layer)	93.8858	1.0695	1.0695	1.2947

Fig. 1. Metrics of the different models used.

It is found that 3 – layer kerasregressor model has the highest at the time of making this project. The accuracies may change depending on the system used. However , A 3-layer neural network based keras regressor is recommended to be used for prediction of hose pricing

III. METHODOLOGY



ACKNOWLEDGMENT

I would like to thank AITS for assigning me this project as a part of my internship and my fellow colleagues for the motivation and enthusiasm to complete this project

REFERENCES

The following resources were referred to while making this project.

[1] https://keras.io/layers/core/

[2] https://www.tensorflow.org/tutorials/keras/basic_regression

[3] https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

[4] “The Complete Reference Python” By Martin C. Brown

[5] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, and Wayan Firdaus Mahmudy, “Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization,” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017

[6] R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher,”real estate value prediction using multivariate regression models” IOP publishing ltd