Assignment 2

Group 8

Hosein Shakib Khankandi (2015478)

Sedigheh Tongou (2102372)

Hamideh Afshari (2104438)

Negar Fooladvand (2107065)

Mohammad Reza Hosseini (1916930)

University Canada West

BUSI 651 Machine Learning Tools and Techniques

Alsaleh Rushdi

November 15th, 2022

Question 1) Develop a Neural Network (NN) prediction model for the sale price. Rationalize the chosen structure of the network.

Answer 1:

A neural network model (file attached) is developed to predict the sale price of houses using the provided dataset. According to the dataset, we have 1460 different recodes of the sale price in millions and four attributes of houses, including the number of garages, garages area, overall quality, and the ground living area.

We use python along with some of its libraries like Pandas, NumPy, and Keras to develop the model. We used 80% (1186 rows) of the data to train the model and 20% (292 rows) of it for assessing the model.

As we have 4 attributes effective in the price of houses, the input layer of the model contains 4 nods. As the predicted price needs to be a continuous positive number, we choose the **Relu** function as the nods' activation function and the **Mean Squared Error (MSE)** as the loss function.

To acquire the best possible model, some observations have been made by changing the number of hidden layers and the number of nods in each layer. The result of these observations is summarized in the following table:

Observation	Number of layers	Layer 1 nodes	Layer 2 nodes	Layer 3 nodes	Layer 4 nodes	Epochs	MAPE
1	4	64	64	64	64	5000	16.08%
2	4	64	64	64	64	1000	16.40%
3	3	64	64	64		1000	16.34%
4	3	32	32	32		1000	20.49%
5	3	128	128	128		1000	16.36%
6	2	64	64			1000	20.97%
7	2	128	128			1000	19.27%

The table shows the MAPE (Mean Absolute Percentage Error) for each observation. The lowest MAPE is observation number 1, with 4 hidden layers and 5000 Epochs. Observation number 3, with three hidden layers, 64 nodes and 1000 Epochs, is the second-best design regarding MAPE. As the result of the observation that number 3 is approximately equal to number 1 and needs far less training time (1000 Epochs rather than 5000 Epochs), we choose this architecture for our model.

Question 2)

What is the sales price forecast if the Garage Car is 2, the Garage area is 600ft2, the overall quality is 7, and the ground living area is 2,200ft2, Comment on the forecasted value.

Answer 2:

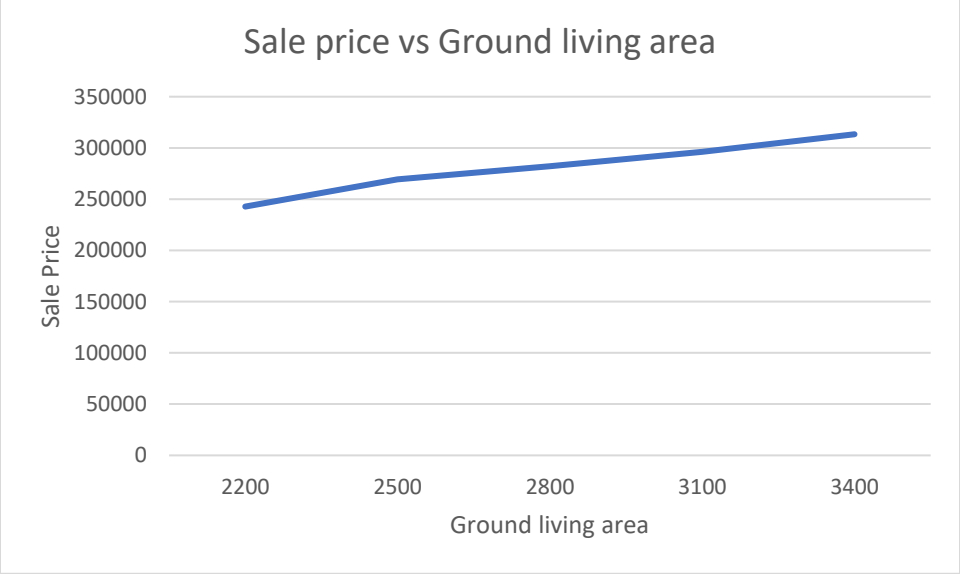
Using the developed model, the price prediction for the house with the given attributes is **242,757 $**. We also filter the dataset to visualize the price of houses in the dataset with similar attributes. We can see that the model predicts the price in the same range as the training dataset.

SalePrice	GarageCars	GarageArea	OverallQual	GrLivArea
180000	1	576	7	2207
244000	0	516	7	2223
222000	0	642	7	2236
225000	1	746	7	2243
236000	2	554	7	2256

Question 3) Simulate the impact of the model variables on the net profit.

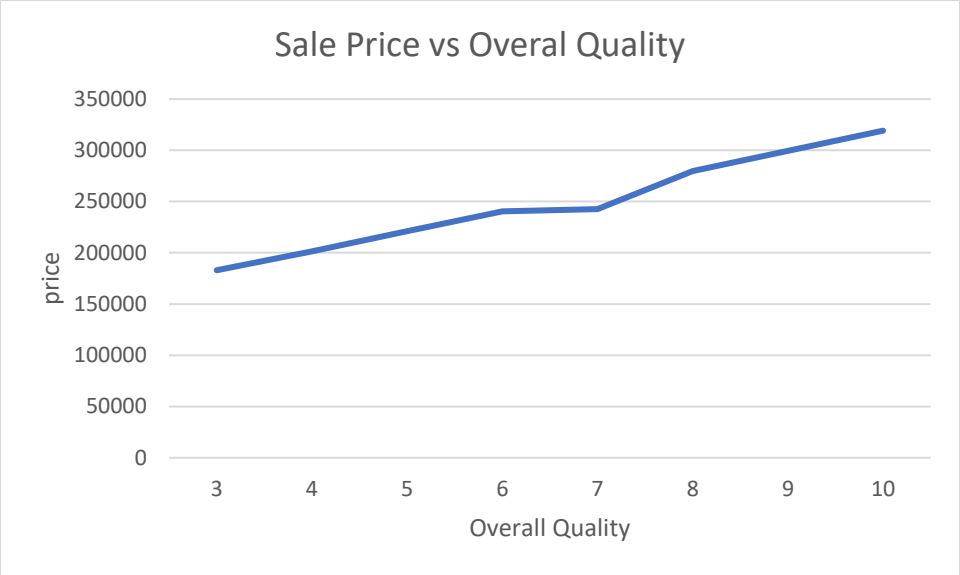
We used the developed model to predict a set of data varied in terms of each variable. In the first simulation, we investigated the influence **of the ground living area** on the predicted price while other attributes were fixed. The following table and figure show that at the chosen condition of the investigation, the price has a relatively linear correlation with the ground living area.

Predictions	GarageCars	GarageArea	OverallQual	GrLivArea	Predicted sale price
1	2	600	7	2200	242757
2	2	600	7	2500	269458
3	2	600	7	2800	282160
4	2	600	7	3100	296099
5	2	600	7	3400	313336



We also investigated the effect of the Overall Quality on the predicted price using the developed model. The result in the following table reflects the relation between the price and the overall quality in the other 3 attributes that are fixed.

Predictions	GarageCars	GarageArea	OverallQual	GrLivArea	Predicted sale price
1	2	600	3	2200	182989
2	2	600	4	2200	201289
3	2	600	5	2200	220938
4	2	600	6	2200	240586
5	2	600	7	2200	242757
6	2	600	8	2200	279883
7	2	600	9	2200	299532
8	2	600	10	2200	319180



Following the same method, the model can help users simulate the effect of changes in any of the attributes on the price of the house.

Question 4)

What are the limitations of the model?

Answer 4:

The model predicts the price based on the dataset it uses to be trained and the variables this dataset gives to the model. The precision of the model is limited due to the quality and the quantity of the dataset it is trained with, and the model will have better prediction with the more extensive dataset to be trained with. The model's predictions are also based on the variables in the dataset. For example, the location of the house, its access to public transformation, the age of the house, and other factors are attributes the model ignores because the training data doesn't provide any information about them.