Predicting House Prices (Keras - Artificial Neural Network)

Table of Contents

1. [Overview](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch1)
2. [Dataset](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch2)
3. [Exploratory Data Analysis](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch3)
4. [Working with Feature Data](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch4)
5. [Feature Engineering](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch5)
6. [Scaling and Train Test Split](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch6)
7. [Creating a Model](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch7)
8. [Training the Model](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch8)
9. [Evaluation on Test Data](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch9)
10. [Predicting on a Brand New House](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#ch10)

## **Overview**

One of the objectives of this notebook is to **show step-by-step how to analyze and visualize the dataset to predict future home prices.** Moreover, we are going to explain most of the concepts used so that you understand why we are using them. In base of features like sqft\_living, bathrooms, bedrooms, view, and others, we are going to build a deep learning model that can predict future price houses.

## **Dataset**[**¶**](https://www.kaggle.com/code/tomasmantero/predicting-house-prices-keras-ann#Dataset)

* This dataset contains house sale prices for King County, which includes Seattle.
* It includes homes sold between May 2014 and May 2015.
* 21 columns. (features)
* 21597 rows.

**Feature Columns**

* **id:** Unique ID for each home sold
* **date:** Date of the home sale
* **price:** Price of each home sold
* **bedrooms:** Number of bedrooms
* **bathrooms:** Number of bathrooms, where .5 accounts for a room with a toilet but no shower
* **sqft\_living:** Square footage of the apartments interior living space
* **sqft\_lot:** Square footage of the land space
* **floors:** Number of floors
* **waterfront:** - A dummy variable for whether the apartment was overlooking the waterfront or not
* **view:** An index from 0 to 4 of how good the view of the property was
* **condition:** - An index from 1 to 5 on the condition of the apartment,
* **grade:** An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.
* **sqft\_above:** The square footage of the interior housing space that is above ground level
* **sqft\_basement:** The square footage of the interior housing space that is below ground level
* **yr\_built:** The year the house was initially built
* **yr\_renovated:** The year of the house’s last renovation
* **zipcode:** What zipcode area the house is in
* **lat:** Lattitude
* **long:** Longitude
* **sqft\_living15:** The square footage of interior housing living space for the nearest 15 neighbors
* **sqft\_lot15:** The square footage of the land lots of the nearest 15 neighbors

## **Evaluation on test data**

### Regression Evaluation Metrics

**Mean Absolute Error** (MAE) is the mean of the absolute value of the errors:

1𝑛∑𝑖=1𝑛|𝑦𝑖−𝑦̂𝑖|

**Mean Squared Error** (MSE) is the mean of the squared errors:

1𝑛∑𝑖=1𝑛(𝑦𝑖−𝑦̂𝑖)2

**Root Mean Squared Error** (RMSE) is the square root of the mean of the squared errors:

1𝑛∑𝑖=1𝑛(𝑦𝑖−𝑦̂𝑖)2‾‾‾‾‾‾‾‾‾‾‾‾‾‾⎷

Comparing these metrics:

* **MAE** is the easiest to understand, because it's the average error.
* **MSE** is more popular than MAE, because MSE "punishes" larger errors, which tends to be useful in the real world.
* **RMSE** is even more popular than MSE, because RMSE is interpretable in the "y" units.