Jaime wang and ziyao Guo

Professor Alan Burstein

CS521 Information Structures with python

12/09/2021

For the final project we choose the topic is house price forecasting. Housing prices are an important indicator of the economic operation, real estate developers and home buyers are paying close attention to the fluctuations in housing prices, and the construction of an effective housing price forecast model is of great significance to the financial market and people's livelihood. In the whole process of model construction, we first preprocess the data, and extract key features to analyze housing prices through visualization and basic model pre-training.

So far, we have been going well. It is very intuitive to see that some basic housing information can be extracted from the folder; then, through tables and charts show the sales price and housing type of comparative data information, which are the key information that customers are concerned about buying a house, such as which house residential style of the highest price, the size of each lot or the development time of the impact on the sales price; the use of linear regression charts shows that the different correlations with sale price are assessed through three aspects: the ground above the living area square feet, basement area and the assessment of the overall material and decoration of the house.

We also mapped the distribution density of house prices. By controlling the parameter distribution of the fit, it is possible to visually evaluate its correspondence with the observed data (black lines are the definite distribution, but the figure shows that this is the nonvalue in the data,

i.e. no fitting is performed). Meantime, we delete a column that contains more than five "none" values. Use get dummies function to process the data. During the modeling process, using kfold divided the data into training and test sets. As an independent data in training, he is not involved in training at all, but he can be used to evaluate the final model. Bayesian regression distribution is a very flexible distribution, with different shape parameters and scale parameters. For example, 1e-06 is the default parameter value for Bayesian Ridge shape parameters inside scikit-learn. This is followed by the Bayesian Regression Distribution, which is a very flexible distribution, with differences in the distribution shapes of different shape parameters and scale parameters, such as the use of 1e-06, which is the default parameter value for the Bayesian Ridge shape parameters within scikit-learn. We also took advantage of the model and got the difference between the price and the original price. The image shows that the test price and the predicted price roughly coincide, indicating the accuracy of our model forecast.