

Housing Price Prediction

Devendra Samatia Vinay Jain
15ucs037@lnmiit.ac.in 15ucs055@lnmiit.ac.in

Divyanshu Jhawar
15ucs040@lnmiit.ac.in

February 13, 2018

1 Problem statement

We are needed to predict the price of house based on given features.

2 Dataset

We took dataset from a github repository[1]. The Dataset have 10 features and 20640 entries. The feratures are:

- longitude
- latitude
- housing median age
- total rooms
- total bedrooms
- population
- households
- median income
- median house value
- ocean proximity

2.1 Issues with Dataset

At first glance we see two problems with the dataset. First one is that are 167 null valued entries for the total bedrooms column and second id that the ocean proximity is not numerical which is not ideal for linear regression. To fix the first problem we replace the null entries with the median of the remaining entries of the column. As 167 is very small in comparision to 20640, it won't create much problem to add null values with median but if we drop the whole column we lose lot if important data. The second problem can be eliminated by using one hot encoder.

2.2 Exploring other features

If we look at the corelationship matrix

References

- [1] ageron. handson-ml.