**3.nl KNN**

We apply KNN method to predict price. The most primitive price prediction algorithm of Airbnb is to use the idea of KNN, which predicts the most appropriate price of a new listing from a set of geographically close listings. Since KNN-regression users Euclidean distance to define the neighbor, only quantitative predictors will be used. We add in all possible predictors by 3 steps, as shown in Table XX. The first group starts with geographical information, as the primitive Airbnb algorithm, and then add in other predictors. Within each group, we conduct cross-validation to determine the best number of neighbors—k.

Table 1 Training and testing RMSE of KNN regression

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Group 1** | |  | **Group 2** | |  | **Group 3** | |
|  | Longitude, latitude | |  | Group 1 + beds, bedrooms, accommodates, bathrooms, amenities | |  | Group 2 + review scores rating, number of reviews, host response rate | |
| **k** | Train RMSE | Test RMSE |  | Train RMSE | Test RMSE |  | Train RMSE | Test RMSE |
| 1 | 0.00 | 1.02 |  | 0.00 | 1.02 |  | 0.00 | 1.02 |
| 5 | 0.54 | 0.86 |  | 0.42 | 0.92 |  | 0.44 | 0.91 |
| 10 | 0.58 | 0.83 |  | 0.46 | 0.90 |  | 0.47 | 0.89 |
| 25 | 0.60 | 0.82 |  | 0.49 | 0.88 |  | 0.50 | 0.87 |
| 50 | 0.61 | 0.81 |  | 0.50 | 0.87 |  | 0.52 | 0.86 |
| 250 | 0.63 | 0.79 |  | 0.53 | 0.84 |  | 0.54 | 0.84 |
| 500 | 0.64 | 0.78 |  | 0.55 | 0.83 |  | 0.56 | 0.82 |
| 840 | 0.65 | 0.77 |  | 0.56 | 0.81 |  | 0.57 | 0.80 |
| 1000 | 0.65 | 0.77 |  | 0.57 | 0.80 |  | 0.58 | 0.80 |
| 3000 | 0.67 | 0.75 |  | 0.61 | 0.77 |  | 0.62 | 0.77 |

Note: All predictors except for longitude and latitude are scaled before training model. Data are randomly divided into training and testing datasets. The best k is searched from a series from 1 to 3,000. Note that in training dataset, the numbers of listings in each city are in the range of 1,264 (Boston) to 11634 (New York City).

**3.nb Regression Tree**

We apply regression tree method to predict prices of Airbnb listings.