Chenyang Yan cy477
Yutong Wang yw2364
Hongyi Nie hn327

ORIE 5250: Final Project Proposal

Overview:

For the second dataset "Expedia", in question four, we defined two types of customers based on whether the customer wants to make an early or a late reservation, but we hope to divide the customer into more segments by using K- means, and K- nearest center, then we use the rest of features to estimate the sensitivity parameters for each type of customers using MLE estimation.

Plan:

We plan to further define different types of customers, using K-means and K-nearest center techniques, to better provide the optimal subsets of hotels to display. We want to compare the results of defining 2,3 and 4 types of customers, then compare the performance and avoid the overfitting of the model.

For different clusters of customers, we will use the mixture of MNL models to estimate the probability of each customer choosing hotels and sensitivity parameters for each type of customer using MLE estimation. We will calculate the revenue for each type of customer to decide the optimal subsets to display. In the meantime, we will also calculate the revenue under a single MLE model. Compare the mixture MNL and the single MNL model result.

Future Work:

After getting the sensitivity parameters for each type of customers, we can customize the ranking for each customer segment to increase revenue.