

ASSIGNMENT-2

Q1. Find the set of cars that have the highest similarity with each other.

Sol. Steps to calculate highest similarity:

1. Read the data.
2. Name header to the data
3. Convert it into a matrix using matrix function
4. Rank the data by creating a function
5. Normalize the data to range [0,1] using

$$z_{if} = \frac{r_{if} - 1}{M_f - 1}$$

6. Computed similarity matrix **sim [1728 X1728]**
7. Find the max value other than diagonal value

Value: 0.6666666666667

Position: sim[1,28], sim[28,1]

Q2. Find the set of cars that have the highest dissimilarity with each other.

Sol. Steps to calculate highest similarity:

1. Same above 5 steps in question 1.
2. Computed dissimilarity matrix **dissim [1728 X1728]**
3. Find the max value other than diagonal value

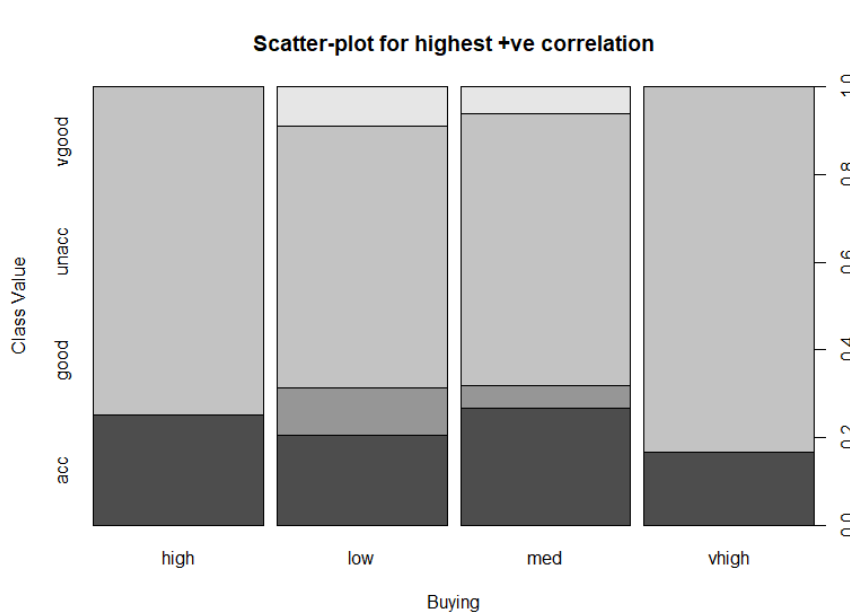
Value: 2.64575131106459

Position: sim[1,1728], sim[1728,1]

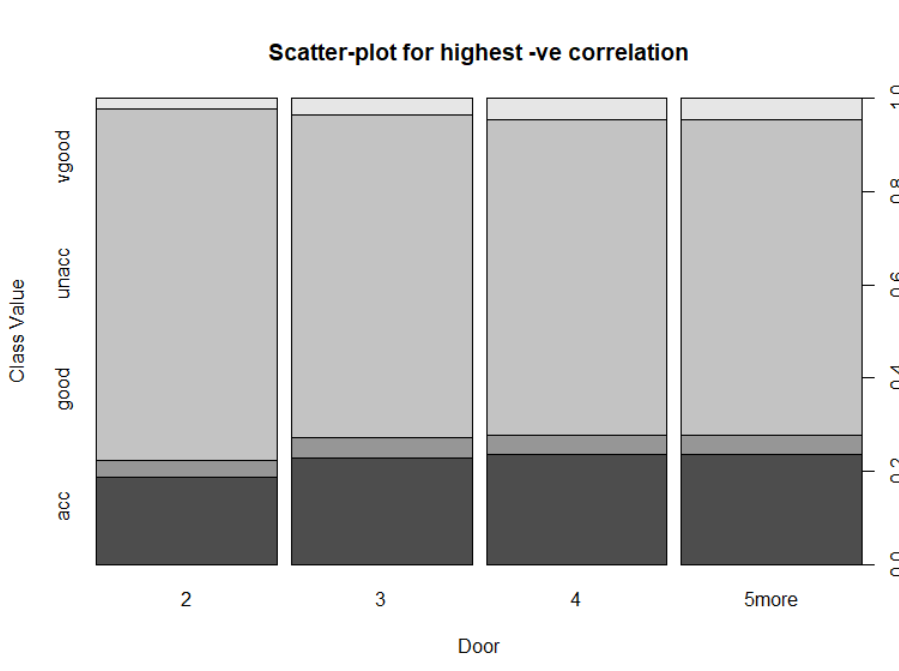
Q3. Find the two categorical attributes that have the highest positive/negative correlation. Draw a scatter plot of these two attributes. Analyze the correlation using the following significance levels:

0.01, 0.05, 0.001

Sol. **Buying and Class** value are the two categorical attribute which have the **highest positive** correlation, as shown below in plot:



Doors and Class value have the **highest negative** correlation, as shown below in plot:



Q4. Present the dissimilarity matrix for the very good cars.

Sol. Steps to present dissimilarity matrix for very good cars:

1. Filter the data for class value= vgood
2. Do the steps 1 to 5 which done in question 1.
3. Compute the dissimilarity matrix.