1.) In this question we are going to design a machine learning problem for credit card fraud detection. You have to predict, based on certain attributes, whether a transaction is fraudulent or legitimate. Answer the following questions.

• 1 a. Is this a classification problem or a regression problem? (2 pts)

Classification Problem because the machine learning problem is show whether a transaction is fraudulent or legitimate. So, in other words it has two results.

• 1 b. Describe at least 4 discrete features of each transaction that is going to help in our problem. (8 pts)

Credit Limit, transaction number, purchase history, purchase cost

• 1 c. Describe at least 3 continuous features of each transaction that is going to help in our problem. (6 pts)

Time of purchase, Location of purchase, Time between purchase

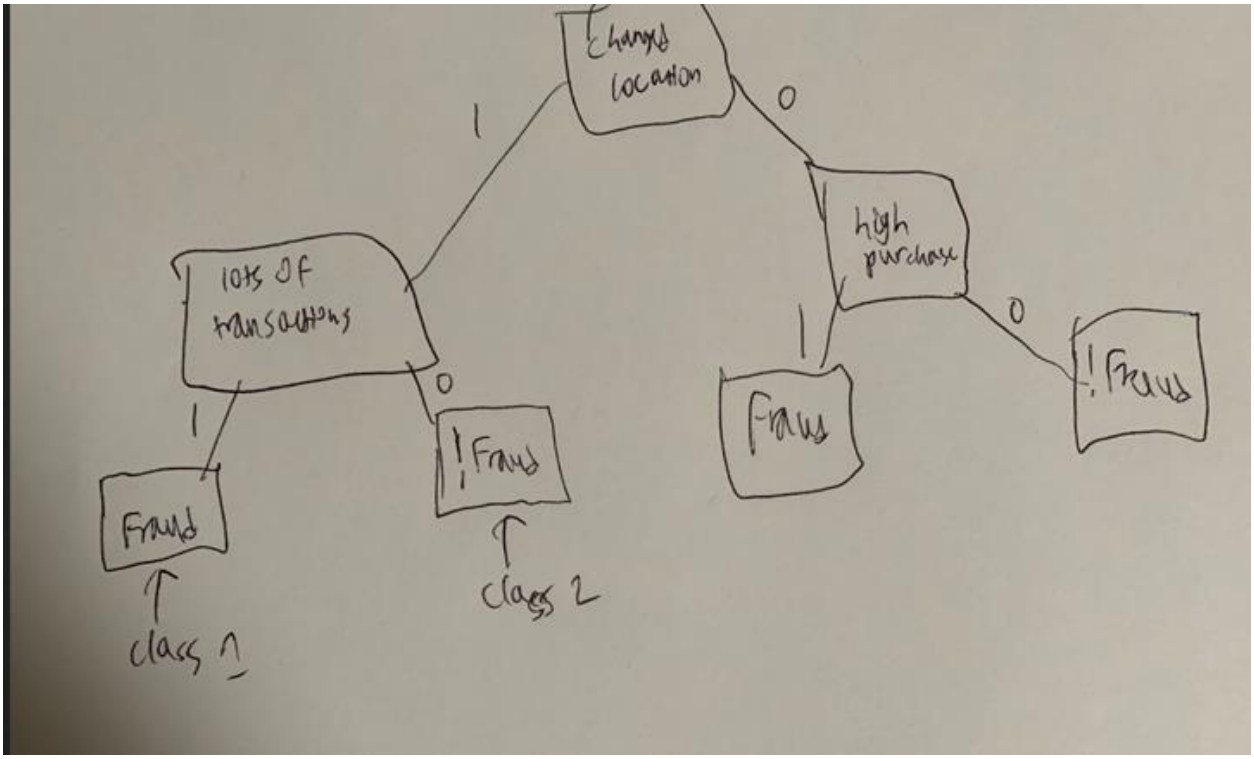
• 1 d. If we were to use an ML algorithm that only takes discrete features as input, what can we do to the continuous features you described in 1 (c) so that they can be used by the ML algorithm? (4 pts)

we can instantiate ranges of possible values, and putting the continuous values into the correlating ranges

• 1 e. How would you know if your model suffers from overfitting? If you are using decision tree in your problem, how would you deal with overfitting? (4 pts)

I would know my model is suffering from overfitting if it performs well on dummy data, but not on actual data. To prevent this I would prune parts of the tree that keep it from growing to it s max depth

• 1 f. In Table 1, observe the data and use your own intuition to draw a decision tree that can classify the data into Class 1 and Class 2. (6 pts)


Changed location
1
0
lots of transactions
high purchase
1
0
1
0
Fraud
!Fraud
Fraud
!Fraud
Class 1
Class 2