Feature preprocessing and generation with respect to models
TOTAL POINTS 5

1.Question 1
Suppose we have a feature with all the values between 0 and 1 except few outliers larger than 1. What can help us to decrease outliers' influence on non-tree models?

(**X**) Apply np.log1p(x) transform to the data

() MinMaxScaler

(**X**) Winsorization

(**X**) Apply rank transform to the features

(**X**) Apply np.sqrt(x) transform to the data

() StandardScaler

1 point

2.Question 2
Suppose we fit a tree-based model. In which cases label encoding can be better to use than one-hot encoding?

(**X**) When we can come up with label encoder, that assigns close labels to similar (in terms of target) categories

(**X**) When categorical feature is ordinal

(**X**) When the number of categorical features in the dataset is huge

2 points

3.Question 3
Suppose we fit a tree-based model on several categorical features. In which cases applying one-hot encoding can be better to use than label-encoding?

() When the feature have only two unique values

(**X**) If target dependence on the label encoded feature is very non-linear, i.e. values that are close to each other in the label encode feature correspond to target values that aren't close.

1 point

4.Question 4
Suppose we have a categorical feature and a linear model. We need to somehow encode this feature. Which of the following statements are true?

() Label encoding is always better than one-hot encoding

() One-hot encoding is always better than label encoding

(**X**) Depending on the dataset either of label encoder or one-hot encoder could be better

1 point

I, Chun-Min Jen, understand that submitting work that isn’t my own may result in permanent failure of this course or deactivation of my Coursera account. Learn more about Coursera’s Honor Code