Feature preprocessing and generation with respect to models
TOTAL POINTS 6

1.Question 1
What type does a feature with values: [‘low’, ‘middle’, ‘high’] most likely have?

() Datetime

(**X**) Ordinal (ordered categorical)

() Categorical

() Coordinates

() Numeric

() Text

1 point

2.Question 2
Suppose you have a dataset X, and a version of X where each feature has been standard scaled.

For which model types training or testing quality can be much different depending on the choice of the dataset?

(**X**) Neural network

(**X**) Nearest neighbours

() GBDT

() Random Forest

(**X**) Linear models

2 points

3.Question 3
Suppose we want to fit a GBDT model to a data with a categorical feature. We need to somehow encode the feature. Which of the following statements are true?

() Label encoding is always better to use 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

4.Question 4
What can be useful to do about missing values?

(**X**) Nothing, but use a model that can deal with them out of the box

() Impute with feature variance

() Apply standard scaler

(**X**) Impute with a feature mean

(**X**) Reconstruct them (for example train a model to predict the missing values)

(**X**) Replace them with a constant (-1/-999/etc.)

(**X**) Remove rows with missing values

2 points

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