

Feature preprocessing and generation with respect to models

Total points 6

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

1 point

- ☐ Numeric
- ☐ Categorical
- ☐ Ordinal (ordered categorical)
- ☐ Coordinates
- ☐ Datetime
- ☐ Text

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

2 points

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

- ☐ Nearest neighbours
- ☐ Linear models
- ☐ Random Forest
- ☐ Neural network
- ☐ GBDT

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?

1 point

- ☐ One-hot encoding is always better than label encoding
- ☐ Label encoding is always better to use than one-hot encoding
- ☐ Depending on the dataset either of label encoder or one-hot encoder could be better

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

2 points

- ☐ Impute with a feature mean
- ☐ Remove rows with missing values
- ☐ Apply standard scaler
- ☐ Reconstruct them (for example train a model to predict the missing values)
- ☐ Replace them with a constant (-1/-999/etc.)
- ☐ Nothing, but use a model that can deal with them out of the box
- ☐ Impute with feature variance

Coursera Honor Code [Learn more](#)

- ☐ I, **Chinmay kumar Das**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit

Save draft