PRACTICE QUIZ • 8 MIN

## Feature preprocessing and generation with respect to models

Submit your assignment — Try again

Receive grade — TO PASS 75% or higher

Grade 100% — View Feedback — We keep your highest score

## 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)
   - ○ Ordinal (ordered categorical)
   - ○ Coordinates
   - ○ Numeric
   - ○ Categorical
   - ○ 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
   - ☐ GBDT
   - ☐ Random Forest
   - ☐ Neural network
   - ☐ Linear models

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

☐ I, **Hong-Yuan Lin**, 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

Save — Submit