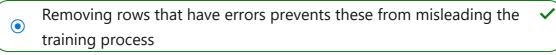✓ 200 XP ▶

Knowledge check

3 minutes

Check your knowledge

1. Why do we clean our data before training?

- ○ Removing rows of data makes our model more powerful

- ○ Cleaning data helps us select features that will be helpful for the model's performance

- ◉ Removing rows that have errors prevents these from misleading the training process ✓

  Correct. Cleaning data helps prevent errors from incomplete or error-prone data points.

2. What kind of data are best encoded with one-hot vectors?

- ○ Ordinal data

- ○ Categorical data with two possible values

- ◉ Categorical data with three or more values ✓

  Correct. One-hot vectors are best used with multiple classes or categories so models can better interpret them.

3. What is a data sample? What is a population?

- ○ A sample is all possible data we care about. A population is the subset of that data which we actually have on hand

- ○ Both population and sample mean data we train our model with.

- ◉ A population is all possible data we care about. A sample is the subset of that data which we actually have on hand. ✓

  Correct. A population is all the possible data we could collect for a data set, and a sample is a portion of the data which we already have.

4. You have a model that isn't performing well. Which of these will definitely **not** help improve its performance?

○   Adding additional samples (rows)

○   Adding a small number of features (columns) that you know are
     related to what the model is trying to predict

⦿   Adding a large number of features that you know aren't related to   ✓
     what the model is trying to predict.

**Correct. Adding additional features that aren't relevant will probably
harm its performance**

Next unit: Summary

Continue   >