200 XP

Knowledge check

3 minutes

Answer the following questions to check your learning.

1. You are using scikit-learn to train a regression model from a dataset of sales data. You want to be able to evaluate the model to ensure it will predict accurately with new data. What should you do?

- ○ Use all of the data to train the model. Then use all of the data to evaluate it
- ○ Train the model using only the feature columns, and then evaluate it using only the label column
- ◉ Split the data randomly into two subsets. Use one subset to train the model, and the other to evaluate it ✓

That is correct. A common way to train and evaluate models is to hold-back an evaluation dataset when training.

2. You have created a model object using the scikit-learn LinearRegression class. What should you do to train the model?

- ○ Call the predict() method of the model object, specifying the training feature and label arrays
- ◉ Call the fit() method of the model object, specifying the training feature and label arrays ✓

That is correct. To train a model, use the fit() method.

- ○ Call the score() method of the model object, specifying the training feature and test feature arrays

3. You train a regression model using scikit-learn. When you evaluate it with test data, you determine that the model achieves an R-squared metric of 0.95. What does this metric tell you about the model?

- ◉ The model explains most of the variance between predicted and actual values. ✓

That is correct. The R-squared metric is a measure of how much of the variance can be explained by the model.

- ○ The model is 95% accurate
- ○ On average, predictions are 0.95 higher than actual values

Module complete:

Unlock achievement