Hide menu

SparkML

- **Video:** SparkML Fundamentals
  5 min
- **Video:** Classification and Regression using Apache Spark
  3 min
- **Video:** SparkML Clustering
  3 min
- **Ungraded App Item:** Obtain IBM Cloud Feature Code and Activate Trial Account
  1h
- **Ungraded Plugin:** Jupyter Notebook for Hands-on Lab: Machine Learning with Apache Spark ML
  20 min
- **Ungraded App Item:** Optional: Hands on Lab: Introduction to SparkML
  1h
- **Reading:** Summary & Highlights
  3 min
- **Practice Quiz:** Practice Quiz: SparkML
  5 questions
- **Quiz:** Graded Quiz: SparkML
  10 questions

Graded Quiz: SparkML

Quiz • 20 min

Submit your assignment

Due Aug 13, 11:59 PM EDT **Attempts** 3 every 8 hours

Receive grade

To Pass 60% or higher

Like Dislike Report an issue

Start assignment

Your grade

–

1. Select the best definition of a machine learning system. **1 point**

- ○ A machine learning *system* trains data models and uses that information to calculate results on the known data.
- ◉ A machine learning *system* applies a specific machine learning *algorithm* to train data models. After training the model, the system infers or "*predicts*" results on previously unseen data.
- ○ A machine learning *system* consists of already trained data models that predict results on previously unseen data.
- ○ A machine learning *system* consists of already trained data models that predict results on known data.

2. Which of the following options are true about Spark ML inbuilt utilities? **1 point**

- ☐ Spark ML utilities help during the intermediate steps of data processing, cleaning, and building models.
- ☑ Spark ML inbuilt utilities includes the Feature module.
- ☑ Spark ML inbuilt utilities includes a linear algebra package.
- ☑ Spark ML inbuilt utilities includes a statistics package.

3. Select the statements that are true about Spark's support for machine learning data sources. **1 point**

- ☑ Supports both feature vector and label column data
- ☐ Images are not a common data source
- ☑ Has standard libraries to support images and LIBSVM data types
- ☑ LIBSVM loads the "libsvm" data files and creates a DataFrame with two columns including the feature vector and label.

4. How do you perform supervised machine learning classification on Apache Spark? **1 point**

- ◉ The Spark ML library provides the spark.ml.classification library for classifications.
- ○ The Spark ML library provides the spark.classification library for classifications
- ○ The Spark ML library provides the spark.ml.regression library for regressions
- ○ The Spark ML library provides the spark.regression.library for regressions

5. Select the statements that are true for classification using Apache Spark. **1 point**

- ☐ Classification is a form of an implicit function approximation where the model predicts real valued outputs for a given input.
- ☐ Classification examples include weather predictions, stock market price predictions, house value estimation, and others.
- ☑ The Spark ML model predicts each object's target category or "class."
- ☑ Producing a prediction from a discrete set of possible outcomes from the task is called classification.

6. Select the statements that are true about regression using Apache Spark ML. **1 point**

- ☑ The predicted value is usually a continuous real number, such as a float or integer
- ☑ Regression is a form of an implicit function approximation where the model predicts real valued outputs for a given input.
- ☐ Examples of regression analysis include predicting a sports tournament winner, heads, or tails on a coin toss, classifying images with a pre-set number of distinct categories
- ☑ Examples of regression analysis include Weather predictions, stock market price predictions, house value estimation, and others.

7. Select the answers that correctly fill in the blank. Unsupervised learning _____. **1 point**

- ☑ Is a subset of machine learning algorithms
- ☐ Requires explicit labels mapped to features
- ☑ Automatically learns patterns and latent spaces in the data
- ☑ Does not require explicit labels mapped to features

8. View the following code samples and place the code in the order needed to perform clustering using Spark ML **1 point**

#1 Perform predictions on test data

```
test_data = spark.read.format("libsvm").load("test_data.txt")

predictions = model.transform(test_data)
```

#2 Create a model and train it

```
kmeans = KMeans().setK(5)

model = kmeans.fit(data)
```

#3 Load data

```
data = spark.read.format("libsvm").load("data.txt")
```

- ○ #2, #3, #1
- ○ #1, #2, #3
- ○ #3, #1, #2
- ◉ #3, #2, #1

9. Select the answer that correctly fills in the blank. The Spark MLlib provides a clustering library located at _____ **1 point**

- ○ (clustering.ml.spark)
- ○ (clustering.spark)
- ◉ (spark.ml.clustering)
- ○ (spark.clustering)

10. Select the clustering algorithms for which Spark MLlib provides functions. **1 point**

- ☑ Latent Dirichlet Allocation
- ☑ Gaussian Mixture Models
- ☐ Early Dirichlet Allocation
- ☑ k-means

Upgrade to submit