

## Module 2 : Introducing Machine Learning

1. Which statement describes machine learning?
  - The scientific study of algorithms and statistical models to perform tasks by using inference instead of instructions
  - The creation of machines that can perform tasks that a human would typically perform
  - The compilation of a large set of rules that cover every possible outcome
  - A set of instructions that describe how to achieve a desired outcome
  
2. Which type of training describes a machine learning application that interacts with its environment and learns to take actions that maximize rewards?
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning
  
3. You are creating a machine learning solution for a call center. The objective of the system is to route customers to the appropriate department, and there are eight possible departments. Which type of machine learning problem does this scenario describe?
  - Regression
  - Multiclass classification
  - Binary classification

Correct

---

This scenario describes a multiclass classification problem because the machine learning system must correctly route the caller to one of multiple possible departments.
  
4. You are working on a machine learning problem that requires the system to respond to changes in the environment to improve performance. Which type of machine learning problem does this scenario describe?
  - Regression
  - Supervised learning
  - Reinforcement learning

5. Which stage of the machine learning (ML) pipeline involves verifying that your data is all of a uniform type?
- Problem formulation
  - Model training
  - Data preparation
6. You are working on a machine model that uses data from multiple countries. The countries are listed by using alphabetical abbreviations. Which stages involve converting these abbreviations to numerical values?
- Data preparation
7. Your model is overfitting if it performs well on the training data, but not on the evaluation data.
- True
10. What are requirements for choosing machine learning as a development methodology?
- A pre-defined machine learning model
  - Large dataset
  - A complex dataset
  - A standard dataset

Correct

Before you decide to develop a machine learning solution, you should have a large dataset with a large number of variables.

## Module 3 : ML with SageMaker

1. Which resources help define a machine learning (ML) problem? (Select TWO.)

- Access to labeled data
- A domain expert to consult
- A traditional coded solution
- Sufficient hardware
- A neural network

2. When preparing data for supervised classification machine learning, which attributes should the data have? (Select TWO.)

- Data should be labeled
- Data should contain only instances of the target
- Anyone in the company should be able to access the data
- Data should be generated randomly by using genetic algorithms
- Data should be representative of production

3. What can you learn by examining the statistics of your data?

- Identifying anomalies in the data
- Verifying that the data is formatted correctly
- Removing outliers
- Filling in missing data

4. You have a preprocessed dataset that's ready for use in training a model. How should you divide your training data?

- Use all the data to train the model.
- Split the data into two equal sets. Use one half for training and the other half for testing.
- Split the data into three sets. Use 80% for training, 10% for testing, and 10% for validation.
- Split the data into two sets. Use 80% for training, and 20% for testing and validation.

You can select between single model and multi-model hosting with Amazon SageMaker.

True

6. What is the purpose of a confusion matrix?

- To plot the labels from the predicted dataset
- To show the true or false positives, along with the true or false negatives
- To show the correlation between two columns in the dataset
- To stratify the classes across training and testing datasets

7. What does a correlation heatmap show?

- The level of correlation between features in a dataset
- The level of correlation between the test and the validation data
- The level of correlation between the predicted and actual values
- The level of correlation between encoded and text data

8. Which of the following file formats does pandas support for data importing? (Select TWO.)

JSON

MS Word

CSV

10. What is the goal of an Amazon SageMaker hyperparameter tuning job?

To optimize the validation metrics for training

To optimize the model parameters to produce the best model

## Module 4 : Forecasting

1. Which patterns are common in time series data? (Select TWO.)

Trends

Seasonal

Which use cases apply to forecasting? (Select TWO.)

Predicting the inventory that's required for items in a warehouse

Predicting if an X-ray image contains an abnormality

Predicting the energy consumption of an office

Predicting the sentiment of a review

Determining if two images are of the same person

3. Which datasets could be used as a time series dataset? (Select TWO.)

Sales data that contains items, purchase dates, and quantities

Web logs that contain IP addresses, pages, and timestamps

4. You have a dataset of temperature readings from a weather station. Temperature readings are logged every 5 minutes. You notice that there are several missing values each day. Which approach could you take? (Select TWO.)

Replace the missing values with zero

Forward fill the missing values

Backward fill the missing values

Use the sum of the temperatures for the day to fill the missing values

Remove the records that have the missing data

5. Which scenarios are examples of appropriate downsampling?  
(Select TWO.)

- Using mean to convert temperature readings every minute to an hourly value
- Using mean to convert sales order information during the day to a daily total
- Using sum to convert sales order information during the day to a daily total
- Using sum to convert temperature readings every minute to an hourly value

6. What are examples of seasonality that you might observe in time series data?  
(Select TWO.)

- Quarterly, yearly
- Spring, summer, fall, winter
- Every two years
- One time sales events
- Hourly

The correct explanation for P10, P50, and P90 predictions in Amazon Forecast is:

P10 indicates that 10% of the time, fewer than the predicted value will be ordered.

P50 indicates that 50% of the time, fewer than the predicted value will be ordered (this is the median prediction).

P90 indicates that 90% of the time, fewer than the predicted value will be ordered.

8. Which datasets are required for generating a retail forecast with Amazon Forecast?

- Item data that includes an item and category
- Item stock information that includes a timestamp, item, and stock quantity
- Item pricing data including a timestamp, item and price
- Time series data that includes a timestamp, item, and quantity

9. You are going to use a dataset to generate a forecast. Which steps would you take to use the data that's available to produce the best model? (Select THREE.)

- Use the train\_test\_split function from scikit-learn to create a training and testing dataset
- Use pandas to split the data by time into a training and testing dataset
- Use the training dataset in Amazon Forecast by specifying a backtest window
- Use the testing dataset to compare predicted values with the actual values
- All of the above

## Module 5: Computer Vision

1. Which are common use cases for computer vision?

Image Analysis

Facial recognition

Home security

All of the above

Correct

Image analysis, facial recognition, and home security are all common use cases for computer vision.

2. What is the location of an object in an image called?

A bounding box

3. Which capabilities are provided by Amazon Rekognition? (Select TWO.)

Searching libraries of images and videos

Adding labels to images

Image manipulation

Facial detection

4. When Amazon Rekognition performs predictions, it also provides a score that indicates the level of confidence in the prediction.

True

False

5. What does Amazon Rekognition do with the results after it completes a video analysis?
- Stores the results in an Amazon Relational Database Service (Amazon RDS) database
  - Starts an AWS Lambda function to notify the owner of the job
  - Publishes the results to an Amazon Simple Notification Service (Amazon SNS) queue
  - Stores the results in Amazon Simple Storage Service (Amazon S3)
6. Which features are part of Amazon Rekognition Custom Labels? (Select TWO.)
- UI for labeling images and defining bounding boxes
  - Automated selection of machine learning algorithms
7. What is the minimum number of images that are required to use automated data labeling by Amazon SageMaker Ground Truth?
- 5000
  - 3000
  - 1500
  - 1250
8. What is a confusion matrix?
- A way to test if your model is working
  - A test to determine the accuracy of a classification model

9. Which types of data are included in an Amazon SageMaker Ground Truth manifest file? (Select THREE.)

Confidence value

File type

Creation date

Class name

Number of images

File size

```
{  
    "source-ref": "s3://bucket/images/image1.jpg",  
    "label": "cat",  
    "label-metadata": {  
        "confidence": 0.95  
    }  
}
```

10. Which of the following are steps for preparing a custom dataset for object detection? (Select TWO.)

Collect images

Feature engineering

Train the model

Generate a confusion matrix

## Module 6 : NLP

1. Which issue is not a major challenge for natural language processing (NLP)?
  - Lack of precision
  - Meaning based on context
  - Multiple dependencies
  - Memory limitations
2. Which tasks are common preprocessing tasks for natural language processing (NLP) applications? (Select TWO.)
  - Removing noise
  - Adjusting for context
  - Removing proper nouns
  - Feature engineering
  - Normalizing similar terms
3. Natural language processing (NLP) systems predate machine learning systems.
  - True
4. Which models are common machine learning models for natural language processing (NLP) applications? (Select TWO.)
  - pandas
  - Bag of words
  - Word tokens
  - Term frequency and inverse document frequency
  - Scikit-learn

5. What is not a text-analysis category?

- Auto-correcting text
- Classifying text
- Discovering similarities in text
- Deriving relationships within text

6. Which capabilities are supported by Amazon Transcribe (Select TWO.)

Change audio output in response to SSML tags.

Convert

Correct

Build s

Amazon Transcribe can recognize recorded voices, convert streaming audio into text, and build subtitles for multiple languages.

Translate

**convert audio to text and build subtitles**

7. How can you change the way Amazon Polly pronounces words?

By slowing down the audio output

By add

Correct

By sen

You can change pronunciation of words by inserting SSML tags.

By imr

8. Which capabilities are a part of Amazon Comprehend? (Select TWO.)

- Translate a document into another language
- Identifying the language used in a document
- Identify images in a document
- Determining the sentiment in a document, such as positive, negative, neutral, or mixed
- Convert text into speech

9. Which of the following AWS services would you use to launch a workflow based on input to an Amazon Lex chatbot?

- Amazon Simple Storage Service
- Amazon Athena
- Amazon Lambda

10. You work for a company that builds applications that are used by a global audience. Which services could help you analyze how your customers use your applications? (Select TWO.)

**amazon translate / amazon comprehend**

## Module 7 : GenAI

1. After defining generative artificial intelligence (generative AI), a company begins brainstorming how they can use the technology for their engineering application. The chief technology officer (CTO) asks for some clarification on how the technology works. Which statement best describes how generative AI works?  
 Generative AI is based on users deploying a model.  
 Generative AI is powered by large machine learning models called foundation models (FMs).  
 With FMs, customers can use the same pretrained model to adapt to multiple tasks  
 Generative AI is based on users labeling data and training a model.  
 Generative AI is governed by a set of rules defined by the user.
2. Amazon CodeWhisperer incorporates security scans for enhanced security in your code. Which types of vulnerabilities does CodeWhisperer scan for? (Select TWO.)  
 AWS security best practices.  
 AWS sustainability best practices  
 Open Web Application Security Project (OWASP) top 10 web application security risks.  
 Reference libraries best practices  
 Compliance best practices
3. What type of artificial intelligence (AI) uses pre-trained large models to create content?  
 Expert systems  
 Deep learning  
 Traditional machine learning (ML)  
 Generative artificial intelligence (AI)

4. Which statement accurately describes what prompt engineering is?

- The process of creating new input features from raw data that help machine learning algorithms better capture the underlying relationships.
- The process of designing and refining the instructions for a language model to generate specific types of output
- The process of preparing data that you use with your machine learning (ML) model.
- The process of preparing a model with additional data to better fit your personal use case.

8. What are some of the benefits of using Amazon CodeWhisperer? (Select TWO.)

- Refactor pre-existing code
- Accelerate coding tasks
- Generate documentation
- Enhance application security
- Optimize pre-existing code

9. Which AWS managed service enables a user to access foundations models?

- Amazon CodeWhisperer
- Amazon SageMaker JumpStart
- AWS Inferentia
- Amazon Bedrock

10. Amazon CodeWhisperer generates code snippets to full functions in real-time based on users's comments and existing code. What is a benefit of CodeWhisperer code generation?

- Bypass repetitive coding tasks