Unit – 3

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

Options:

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

✅ Correct Answers:

- Access to labeled data
- A domain expert to consult

---

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

Options:

- 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 ✅

✅ Correct Answers:

- Data should be labeled
- Data should be representative of production

---

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

Options:

- Identifying anomalies in the data ✅

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

✅ Correct Answer:

- Identifying anomalies in the data

---

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

Options:

- 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

✅ Correct Answer:

- Split the data into three sets. Use 80% for training, 10% for testing, and 10% for validation

---

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

Options:

- True ✅
- False

✅ Correct Answer:

- True

---

6. What is the purpose of a confusion matrix?

Options:

- 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

✅ Correct Answer:

- To show the true or false positives, along with the true or false negatives

---

7. What does a correlation heatmap show?

Options:

- 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

✅ Correct Answer:

- The level of correlation between features in a dataset

---

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

Options:

- JSON ✅
- MS Word
- CSV ✅
- Binary files
- PDF

✅ Correct Answers:

- JSON
- CSV

---

9. Which Amazon service can you use to deploy machine learning instances and run Jupyter Notebooks?

Options:

- Amazon Comprehend
- Amazon SageMaker ✅
- Amazon Polly
- Amazon Lex

✅ Correct Answer:

- Amazon SageMaker

---

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

Options:

- To optimize the validation metrics for training
- To optimize the model parameters to produce the best model ✅
- To optimize the data inputs to produce the fastest prediction
- To optimize the algorithm choice to produce the best model

✅ Correct Answer (as expected by exam):

- To optimize the model parameters to produce the best model

Unit – 4

Which use cases apply forecasting? (Select TWO.)

✓ Predicting the inventory that's required for items in a warehouse ✓ Predicting the energy consumption of an office

Why? Forecasting means predicting future numeric values over time → inventory & energy usage.

Why K-Means is a good choice for grouping customers?

✓ Because customers naturally form numeric clusters based on similarities

Why? K-Means groups numerical data (spending, visits) into clusters.

Common patterns in time series? (Select TWO.)

✓ Trends ✓ Seasonal

Why?

Trend = long-term increase/decrease

Seasonal = repeating pattern (daily, weekly, yearly)

What causes K-Means to converge to a poor local minimum?

✓ Initial centroids are placed in sparse or outlier regions

Why? Bad starting points → bad clusters.

Which datasets are time series? (Select TWO.)

✓ Sales data that contains items, purchase dates, and quantities ✓ Web logs that contain IP addresses, pages, and timestamps

Why? Time series requires timestamp + changing values.

Updated centroid for points (1,1), (2,2), (3,1)

✓ (2, 1.33)

Why? Mean X = (1+2+3)/3 = 2 Mean Y = (1+2+1)/3 = 1.33

Required items for generating retail forecast in SageMaker Canvas

✓ Time series data that includes a timestamp, item and quantity

Why? Forecasting always needs timestamp + target value.

Which robot is MOST likely to change cluster?

✓ (2.1, 0.9)

Why? New centroid moved from (1,1) → (2,1). The point closest to new centroid = (2.1, 0.9).

Why sparse cluster centroid moves toward dense cluster?

✓ K-Means gives equal weight to all clusters regardless of their size

Why? Dense cluster has more points → pulls centroid toward itself.

New value of centroid C_1 after 1 iteration

✓ 4

Why? Cluster C_1 gets: 2, 4, 6 New centroid = (2 + 4 + 6) / 3 = 4

Unit – 5

1. Which are common use cases for computer vision?

Options:

- Image Analysis
- Facial recognition
- Home security
- All of the above ✅

✅ **Correct Answer:**
All of the above

---

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

Options:

- A bounding box ✅
- An object box
- Object coordinates
- Object location

✅ **Correct Answer:**
A bounding box

---

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

Options:

- Searching libraries of images and videos ✅
- Adding labels to images
- Image manipulation
- Facial detection ✅
- Video editing

✅ **Correct Answers:**

- **Searching libraries of images and videos**
- **Facial detection**

---

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

Options:

- True ✅
- False

✅ **Correct Answer:**
True

---

5. What does Amazon Rekognition do with the results after it completes a video analysis?

Options:

- Stores the results in an 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) ✅
- Stores the results in Amazon S3

✅ **Correct Answer:**
Publishes the results to an Amazon Simple Notification Service (Amazon SNS)

---

6. Which features are part of Amazon Rekognition Custom Labels? (Select TWO.)

Options:

- UI for labeling images and defining bounding boxes ✅
- Automated selection of machine learning algorithms ✅
- Retrieval of text from an image
- Facial analysis
- Identification of celebrities

✅ **Correct Answers:**

- **UI for labeling images and defining bounding boxes**
- **Automated selection of machine learning algorithms**

---

7. What is the minimum number of images required to use automated data labeling by Amazon SageMaker Ground Truth?

Options:

- 5000
- 3000
- 1500
- 1250 ✅

✅ **Correct Answer:**
1250

---

8. What is a confusion matrix?

Options:

- A way to test if your model is working
- A test to determine the accuracy of a classification model ✅
- A special output from Amazon Rekognition Custom Labels
- A way to validate a linear regression model

✅ **Correct Answer:**
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.)

Options:

- Confidence value ✅
- File type
- Creation date ✅
- Class name ✅
- Number of images
- File size

✅ **Correct Answers:**

- **Confidence value**
- **Creation date**
- **Class name**

---

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

Options:

- Collect images ✅
- Feature engineering
- Train the model ✅
- Generate a confusion matrix

✅ **Correct Answers:**

- **Collect images**
- **Train the model**