**AWS Certified Machine Learning** - **Specialty Practice Questions**

**Requirement**: Create & Share 10 ML Specialty practice questions. These are for “**Section Tests**” to be placed below the existing practice:

**Topic**

* Questions around core machine learning concepts such as Pearson coefficient and Bayesian classification.

**Delivery Timeline**: July 6

**Question Response Types**

There are two types of questions:

* Multiple Choice Single Response – **1** correct answer **3** incorrect responses (distractors).
* Multiple Choice Multiple Response – **2** or more correct answers out of **5** or more options.

**Important Note**

* Do write Question Number for quick identification. Q# 1, Q# 2 …. & so on.
* Please mention Domain (based on ML Specialty exam blueprint), Topic & Sub-Topic (If Applicable) with every question.
* Note that we’re expecting standard scenario based questions & NOT straight-forward definition kind of questions.
* The options should not have any obviously incorrect option. We need to word the options so that all of them should appear correct for the students, but a subtle point should mark the correct answer without any ambiguity. So, one answer should be the best choice without any doubt.
* The answer / explanation section should contain explanations on why the answer is correct and others are incorrect. It should also contain the relevant resource link (for details) preferably from AWS documentation.
  + Example
    - Option A is incorrect because..
    - Option B is CORRECT because…
    - Option C is incorrect because..
    - Option D is incorrect because..
* Try to balance the domains based on weightage % defined in the exam blueprint.
* Any AWS service or feature must be approximately 6 months old to figure out in Practice Tests. Put a note in the explanation for any latest service or feature that might be on the borderline of appearing in the real exam.
* **Plagiarism** in any form - Question or in Explanation will be **rejected.** Questions & Explanations should reflect your own professional experience & AWS skills. Author’s who indulge in plagiarism will be **blacklisted** & dropped from our author’s list.
* The ownership of the questions once approved & published on Whizlabs LMS platform, lies solely with Whizlabs Software Pvt. Ltd. You can’t share or publish it elsewhere in any circumstances.

**Sample Format of** **Questions**

----------------------------------------------------------------------------------------------------------------------------

**Question​ ​:​** #

**Main​ ​Topic​ ​:​** < >

**Sub​ ​Topic​ ​:​** [optional]

**Domain:** < >

**Question text**:

<Scenario based. Should be clear in terms of requirements. No ambiguity. No duplicate options. In case of multiple answers, at the end, you should include number of expected answers. e.g. (Select TWO) or (Select THREE) etc. For single answers this is NOT required>

1. Option A...
2. Option B...
3. Option C...
4. Option D...

**Answer:** A and C

**Explanation:**

**Option A is CORRECT because...**

**Option B is incorrect because...**

**Option C is CORRECT because...**

**Option D is incorrect because...**

[Insert the explanation in clear and lucid language here.]

**Diagram:** [Optional] [Insert the architectural or conceptual diagram here.]

**Reference:** [Insert the references here - which may include links to AWS Documentation, Blog, re:Invent video, Authority YouTube video].

----------------------------------------------------------------------------------------------------------------------------

**ML Specialty has 4 Domains**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Name of the Domain** | **Weight** | **Estimated No. of Questions**  (out of 65 As per weightage %) |
| 1 | Data Engineering | 20% | 13 |
| 2 | Exploratory Data Analysis | 24% | 16 |
| 3 | Modeling | 36% | 23 |
| 4 | ML Implementation and Operations | 20% | 13 |

--------------------------------------Question Section Starts-----------------------------------------------------

Question: 1

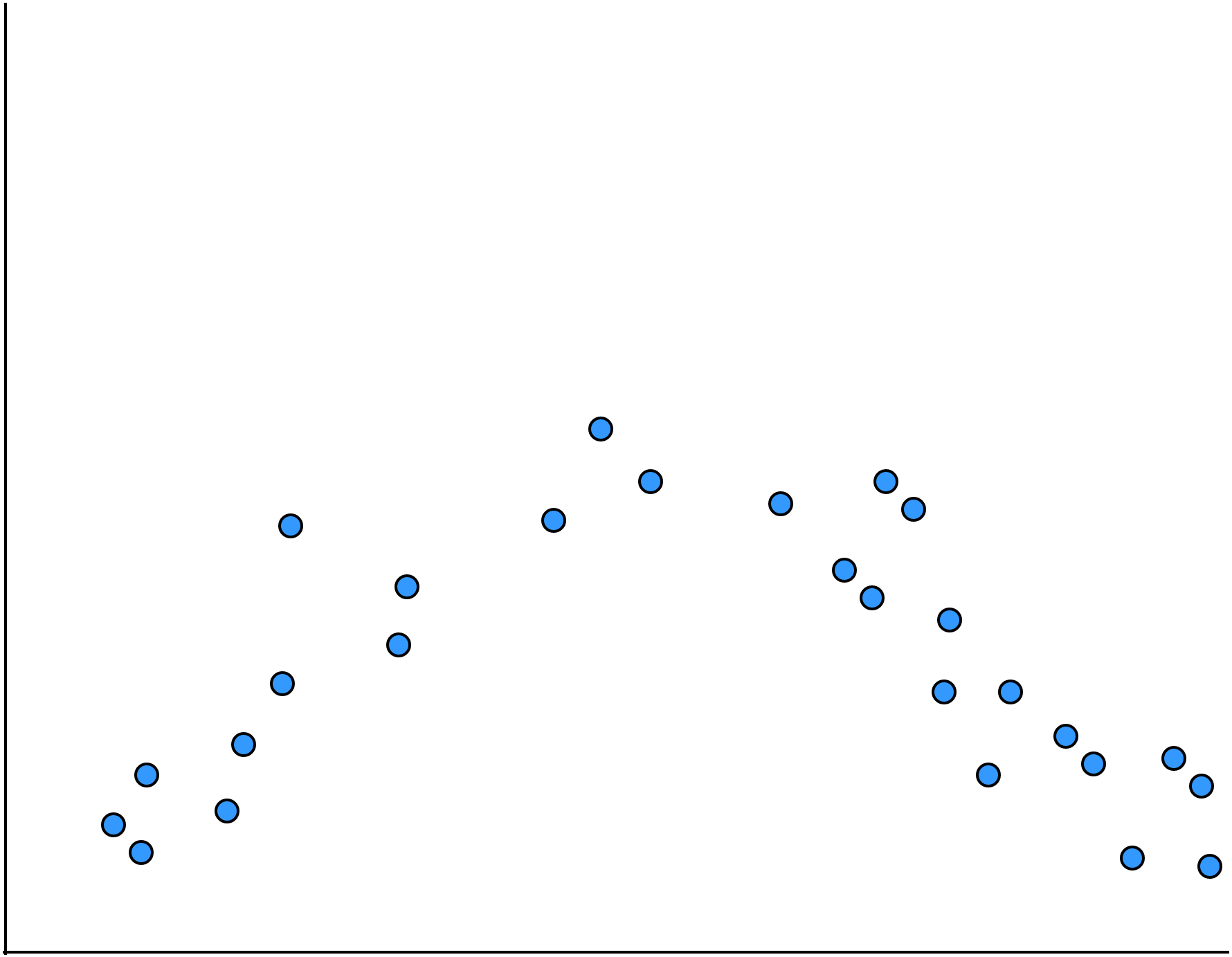
**Main​ ​Topic​ ​:​** Machine Learning

**Sub​ ​Topic​ ​:​ Analyze and visualize data for machine learning**

**Domain:** Exploratory Data Analysis

**Question text**:

You are a machine learning specialist working for an oil refinery company. Your team is working on a machine learning problem where you need to determine the relationship between oil well depth and oil well production. In order to select the appropriate machine learning model to use to attempt to solve the oil well production problem, you need to gain a better understanding of your data. For example, what is the correlation between your oil well depth data and your oil well production data?

When you examine your data visually using the Python matplotlib library, you find that your data has what looks like a non-Gaussian distribution of oil well depth and oil well production:  
[](https://www.draw.io/?page-id=kvxQOwqLXI_e5wS3bV51&scale=auto#G1J2dEHLgrH_-3FTQjtKVSFbzoDf21m9qY)

Which correlation coefficient would you use to best summarize the strength of the correlation between your oil well depth and oil well production?

1. Covariance correlation coefficient
2. Pearson’s correlation coefficient
3. Spearman’s correlation coefficient
4. Polychoric correlation coefficient

**Answer:** C

**Explanation:**

Option A is incorrect. Covariance is used when you have a Gaussian relationship between your variables.

Option B is incorrect. Pearson’s correlation coefficient is also used when you have a Gaussian relationship between your variables.

Option C is correct. Spearman’s correlation coefficient is used when you have a non-Gaussian relationship between your variables.

Option D is incorrect. The polychoric is used to understand the relationship of variables gathered via surveys such as personality tests and surveys that use rating scales.

**Reference:**

Please see the Machine Learning Mastery page titled **How to Calculate Correlation Between Variables in Python** (<https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/>), the Wikipedia page titled **Correlation coefficient** (<https://en.wikipedia.org/wiki/Correlation_coefficient>), the the Wikipedia page titled **Polychoric correlation** (<https://en.wikipedia.org/wiki/Polychoric_correlation>), and the Medium article titled **What are Covariance and Correlation coefficients and their significance?** (<https://towardsdatascience.com/covariance-and-correlation-321fdacab168>)

Question: 2

**Main​ ​Topic​ ​:​** Machine Learning

**Sub​ ​Topic​ ​:​ Analyze and visualize data for machine learning**

**Domain:** Exploratory Data Analysis

**Question text**:

You are a machine learning specialist working for a clothing manufacturer. You have been tasked with building a machine learning model to determine the return on investment (ROI) for advertising a specific clothing line on social media based on the labeled data of past social media campaigns for similar clothing lines.

You decide to run a Pearson correlation coefficient to better understand your data correlation. When you calculate your Pearson’s correlation coefficient of social media advertising ROI you get a value of 0.35. What conclusions can you draw from this result?

1. There is an effective relationship between your past social media advertising and corresponding campaign ROI.
2. There is an unfavorable relationship between your past social media advertising and corresponding campaign ROI.
3. There is no correlation between past social media advertising and the associated ROI.
4. You cannot declare a notable correlation with confidence based on the resulting coefficient.

**Answer:** D

**Explanation:**

Option A is incorrect. Your coefficient value is not high enough to indicate a positive relationship. For a Pearson’s correlation coefficient to indicate a notable correlation, the coefficient value should be above 0.5 for a positive correlation, or below -0.5 for a negative correlation. Your score is 0.35, which falls into the indeterminate range.

Option B is incorrect. Your coefficient value is not low enough to indicate a negative relationship. For a Pearson’s correlation coefficient to indicate a notable correlation, the coefficient value should be above 0.5 for a positive correlation, or below -0.5 for a negative correlation. Your score is 0.35, which falls into the indeterminate range.

Option C is incorrect. A coefficient value of 0 or close to 0 indicates no correlation. Your value of 0.35 is not close enough to 0 to indicate no correlation.

Option D is correct. Your coefficient falls into the indeterminate range. For a Pearson’s correlation coefficient to indicate a notable correlation, the coefficient value should be above 0.5 for a positive correlation, or below -0.5 for a negative correlation. Your score is 0.35, which falls into the indeterminate range.

**Reference:**

Please see the Machine Learning Mastery page titled **How to Calculate Correlation Between Variables in Python** (<https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/>), the Wikipedia page titled **Correlation coefficient** (<https://en.wikipedia.org/wiki/Correlation_coefficient>), and the Medium article titled **What are Covariance and Correlation coefficients and their significance?** (<https://towardsdatascience.com/covariance-and-correlation-321fdacab168>)