



**Free Questions for [AIP-210](#) by [ebraindumps](#)**

**Shared by [Levy](#) on [24-05-2024](#)**

**For More Free Questions and Preparation Resources**

**[Check the Links on Last Page](#)**

# Question 1

---

**Question Type:** MultipleChoice

---

Which of the following is TRUE about SVM models?

**Options:**

---

- A- They can be used only for classification.
- B- They can be used only for regression.
- C- They can take the feature space into higher dimensions to solve the problem.
- D- They use the sigmoid function to classify the data points.

**Answer:**

---

C

**Explanation:**

---

SVM models can use kernel functions to map the input data into higher-dimensional feature spaces, where linear separation is possible. This allows SVM models to handle non-linear problems effectively. Reference: CertNexus Certified Artificial Intelligence Practitioner, Support vector machine - Wikipedia

## Question 2

---

**Question Type:** MultipleChoice

---

Which of the following can take a question in natural language and return a precise answer to the question?

**Options:**

---

- A- Databricks
- B- IBM Watson
- C- Pandas
- D- Spark ML

**Answer:**

---

B

**Explanation:**

---

IBM Watson is an AI technology that can take a question in natural language and return a precise answer to the question. IBM Watson is a cognitive computing system that can understand natural language, generate hypotheses, and provide evidence-based answers. IBM Watson can be applied to various domains and industries, such as healthcare, education, finance, or law.

## Question 3

---

**Question Type:** MultipleChoice

---

Which of the following approaches is best if a limited portion of your training data is labeled?

**Options:**

---

- A- Dimensionality reduction
- B- Probabilistic clustering
- C- Reinforcement learning
- D- Semi-supervised learning

**Answer:**

---

D

### **Explanation:**

---

Semi-supervised learning is an approach that is best if a limited portion of your training data is labeled. Semi-supervised learning is a type of machine learning that uses both labeled and unlabeled data to train a model. Semi-supervised learning can leverage the large amount of unlabeled data that is easier and cheaper to obtain and use it to improve the model's performance. Semi-supervised learning can use various techniques, such as self-training, co-training, or generative models, to incorporate unlabeled data into the learning process.

## **Question 4**

---

### **Question Type: MultipleChoice**

---

You have a dataset with many features that you are using to classify a dependent variable. Because the sample size is small, you are worried about overfitting. Which algorithm is ideal to prevent overfitting?

### **Options:**

---

**A-** Decision tree

**B-** Logistic regression

**C-** Random forest

**D-** XGBoost

**Answer:**

---

C

**Explanation:**

---

Random forest is an algorithm that is ideal to prevent overfitting when using a dataset with many features and a small sample size. Random forest is an ensemble learning method that combines multiple decision trees to create a more robust and accurate model. Random forest can prevent overfitting by introducing randomness and diversity into the model, such as by using bootstrap sampling (sampling with replacement) to create different subsets of data for each tree, or by using feature selection (choosing a random subset of features) to split each node in a tree.

## Question 5

---

**Question Type:** MultipleChoice

---

Which of the following is the correct definition of the quality criteria that describes completeness?

### Options:

---

- A- The degree to which all required measures are known.
- B- The degree to which a set of measures are equivalent across systems.
- C- The degree to which a set of measures are specified using the same units of measure in all systems.
- D- The degree to which the measures conform to defined business rules or constraints.

### Answer:

---

A

### Explanation:

---

Completeness is a quality criterion that describes the degree to which all required measures are known. Completeness can help assess the coverage and availability of data for a given purpose or analysis. Completeness can be measured by comparing the actual number of measures with the expected number of measures, or by identifying and counting any missing, null, or unknown values in the data.

## Question 6

---

**Question Type:** MultipleChoice

---

In which of the following scenarios is lasso regression preferable over ridge regression?

**Options:**

---

- A- The number of features is much larger than the sample size.
- B- There are many features with no association with the dependent variable.
- C- There is high collinearity among some of the features associated with the dependent variable.
- D- The sample size is much larger than the number of features.

**Answer:**

---

B

**Explanation:**

---

Lasso regression is a type of linear regression that adds a regularization term to the loss function to reduce overfitting and improve generalization. Lasso regression uses an L1 norm as the regularization term, which is the sum of the absolute values of the coefficients. Lasso regression can shrink some of the coefficients to zero, which effectively eliminates some of the features from the model. Lasso regression is preferable over ridge regression when there are many features with no association with the dependent variable, as it can perform feature selection and reduce the complexity and noise of the model.



## Question 7

---

**Question Type:** MultipleChoice

---

Which of the following is the primary purpose of hyperparameter optimization?

### Options:

---

- A- Controls the learning process of a given algorithm
- B- Makes models easier to explain to business stakeholders
- C- Improves model interpretability
- D- Increases recall over precision

### Answer:

---

A

### Explanation:

---

Hyperparameter optimization is the process of finding the optimal values for hyperparameters that control the learning process of a given algorithm. Hyperparameters are parameters that are not learned by the algorithm but are set by the user before training.

Hyperparameters can affect the performance and behavior of the algorithm, such as its speed, accuracy, complexity, or generalization.

Hyperparameter optimization can help improve the efficiency and effectiveness of the algorithm by tuning its hyperparameters to achieve the best results.

## Question 8

---

**Question Type:** MultipleChoice

---

Which two encodes can be used to transform categories data into numerical features? (Select two.)

### Options:

---

- A- Count Encoder
- B- Log Encoder
- C- Mean Encoder
- D- Median Encoder
- E- One-Hot Encoder

### Answer:

---

C, E

### **Explanation:**

---

Encoding is a technique that transforms categorical data into numerical features that can be used by machine learning models.

Categorical data are data that have a finite number of possible values or categories, such as gender, color, or country. Encoding can help convert categorical data into a format that is suitable and understandable for machine learning models. Some of the encoding methods that can be used to transform categorical data into numerical features are:

Mean Encoder: Mean encoder is a method that replaces each category with the mean value of the target variable for that category.

Mean encoder can capture the relationship between the category and the target variable, but it may cause overfitting or multicollinearity problems.

One-Hot Encoder: One-hot encoder is a method that creates a binary vector for each category, where only one element has a value of 1 (the hot bit) and the rest have a value of 0. One-hot encoder can create distinct and orthogonal vectors for each category, but it may increase the dimensionality and sparsity of the data.

## **Question 9**

---

**Question Type:** MultipleChoice

---

An HR solutions firm is developing software for staffing agencies that uses machine learning.

The team uses training data to teach the algorithm and discovers that it generates lower employability scores for women. Also, it predicts that women, especially with children, are less likely to get a high-paying job.

Which type of bias has been discovered?

**Options:**

---

- A- Automation
- B- Emergent
- C- Preexisting
- D- Technical

**Answer:**

---

C

**Explanation:**

---

Preexisting bias is a type of bias that originates from historical or social contexts, such as stereotypes, prejudices, or discriminations. Preexisting bias can affect the data or the algorithm used for machine learning, as well as the outcomes or decisions made by machine learning. Preexisting bias can cause unfair or harmful impacts on certain groups or individuals based on their attributes, such as gender, race, age, or disability<sup>3</sup>. In this case, the software that uses machine learning generates lower employability scores for women and predicts that women, especially with children, are less likely to get a high-paying job. This indicates that the software has preexisting bias

against women, which may reflect the historical or social inequalities or expectations in the labor market.

## Question 10

---

**Question Type:** MultipleChoice

---

We are using the k-nearest neighbors algorithm to classify the new data points. The features are on different scales.

Which method can help us to solve this problem?

**Options:**

---

**A-** Log transformation

**B-** Normalization

**C-** Square-root transformation

**D-** Standardization

**Answer:**

---

B

**Explanation:**

---

Normalization is a method that can help us to solve the problem of features being on different scales when using the k-nearest neighbors algorithm. Normalization is a technique that rescales the values of features to a common range, such as  $[0, 1]$  or  $[-1, 1]$ . Normalization can help reduce the influence or dominance of some features over others, as well as improve the accuracy and performance of the algorithm.

**To Get Premium Files for AIP-210 Visit**

**<https://www.p2pexams.com/products/aip-210>**

**For More Free Questions Visit**

**<https://www.p2pexams.com/certnexus/pdf/aip-210>**

