Sample Exam

Edition 202503

Content

Introduction

This is the EXIN BCS Machine Learning Award (AIMLA.EN) sample exam. The Rules and Regulations for EXIN's examinations apply to this exam.

This exam consists of 18 multiple-choice questions. Each multiple-choice question has a number of possible answers, of which only one is correct, unless otherwise stated.

The maximum number of points that can be obtained for this exam is 20. For 18 questions, each correct answer is worth 1 point. For 2 questions, the correct answer is worth 2 points. You need 13 points or more to pass the exam.

The time allowed for this exam is 30 minutes.

Good luck!

Sample exam

1 / 18
What is machine learning?

A) The ability to learn from a machine
B) The use of computer algorithms to capture human expertise
C) The use of computer algorithms to derive insight from data
D) The use of learning theory to make sense of data

2 / 18
Which of the following is an example of an object-oriented language?

A) CSS
B) Python
C) R
D) SQL

3 / 18
Which of the following mathematical principles can **best** be used to measure uncertainty to determine an outcome?

A) Calculus
B) Linear algebra
C) Probability
D) Statistics

4 / 18
Machine learning can be used to sort unlabeled data into groups.

What is this known as?

A) Classification
B) Clustering
C) Grouping
D) Prediction

5 / 18
Which **two** of the following languages are commonly used in machine learning?

Please remember to choose 2 answers.

A) CSS
B) MATLAB
C) Python
D) SQL

6 / 18
What is required in order to train a machine learning model?

A) Neural network
B) Teacher
C) Test data
D) Visualization tools

7 / 18
Which of the following **two** frameworks can be used to develop machine leaning models?

Please remember to choose 2 answers.

A) Google analytics
B) Minecraft
C) Scikit-Learn
D) TensorFlow

8 / 18
Satpal has been developing an application (App) that can be used to order food from different restaurants and have it delivered straight to your home. As part of the functionality, he has built in a machine learning model that uses regression to provide the user with an 'expected delivery time' for the food based on the time of day, the distance between the restaurant to the intended location, and the average delivery time.

The algorithm being used has been configured to compare the two variables 'time of day' and 'average delivery time' in order to make its prediction.

When testing the App, he has found that the predicted 'expected delivery time' seems incredibly long based on his location to the restaurant.

What is probably the issue?

A) The average delivery time
B) The average recorded speed of the driver
C) The choice of algorithm
D) The variables being compared in the data

9 / 18
Roisin is training her machine learning model using unlabeled data and no training data.

What type of type of approach would she use?

A) Reinforcement learning
B) Semi-supervised learning
C) Supervised learning
D) Unsupervised learning

10 / 18
Which **two** of the following problems can be solved through classification?

Please remember to choose 2 answers.

A) Grouping sets of unlabeled data to identify different customer segments
B) Identifying an image based on specific features in the data
C) Making predictions on the number of cases of a virus in a particular area
D) Sorting emails into 'received' and 'spam'

11 / 18
Which of the following is open-source software developed by Google to design machine learning models?

A) Google analytics
B) Scikit-Learn
C) TensorFlow
D) Weka

12 / 18
Which type of algorithm is commonly used within deep learning?

A) K-nearest neighbors
B) Linear regression
C) Naïve Bayes
D) Neural network

13 / 18
What is supervised learning?

A) Training a machine learning model through reinforcement
B) Training a machine learning model using unlabeled data
C) Training a machine learning model where the desired output is known
D) Training a machine learning model where the output is unknown

14 / 18
Dale is wishing to develop an application of machine learning that is able to sort different types of user requests that have been manually input into a system. Due to the volume of user requests it is often time-consuming for an individual to read through each request. It is not always easy to quantify or prioritize requests based on how many times the same type of request is made.

Dale has observed that there are specific words that regularly feature in certain types of user requests that could be used to identify them. He would therefore like a machine learning model to read through each request and sort them into defined categories based on whether these specific words feature.

What type of approach should he use to solve the problem?

A) Classification
B) Clustering
C) Regression
D) Reinforcement

15 / 18
When would you use unsupervised learning?

A) When our algorithm does not require our input
B) When we are too busy to oversee the process
C) When we have sets of labeled data
D) When we have sets of unlabeled data

16 / 18
Which of the following algorithms can be used within clustering?

A) K-means
B) K-nearest neighbors
C) Linear regression
D) Logistic regression

17 / 18
What is the purpose of data pre-processing?

Please remember to choose 2 answers.

A) To clean the data to ensure it is suitable for training a machine learning model
B) To identify features or target values in the data that can be used to create training data
C) To identify the types of data required to solve the problem
D) To present the data using a graph or chart

18 / 18
What is the correct order of steps in the machine learning process?

A) 1. Problem identification
2. Data pre-processing
3. Data selection
4. Testing
5. Training

B) 1. Problem identification
2. Data pre-processing
3. Data selection
4. Training
5. Testing

C) 1. Problem identification
2. Data selection
3. Data pre-processing
4. Training
5. Testing

D) 1. Problem identification
2. Training
3. Data selection
4. Data pre-processing
5. Testing

Answer key

1 / 18
What is machine learning?

A) The ability to learn from a machine
B) The use of computer algorithms to capture human expertise
C) The use of computer algorithms to derive insight from data
D) The use of learning theory to make sense of data

A) Incorrect.
B) Incorrect.
C) Correct. Using algorithms, machine learning applications can perform tasks such as classification, image recognition and forecasting through the analysis of data.
D) Incorrect.

2 / 18
Which of the following is an example of an object-oriented language?

A) CSS
B) Python
C) R
D) SQL

A) Incorrect.
B) Correct. An object-oriented language is one which features objects that contain data fields and code. This type of language is commonly used in machine learning where developers wish to manipulate the objects (i.e. the data) rather than manipulate the logic, where large and complex problems cannot be solved through simple logic. Languages like “R” are functional programming languages, although they support object-oriented programming.
C) Incorrect.
D) Incorrect.

3 / 18
Which of the following mathematical principles can **best** be used to measure uncertainty to determine an outcome?

A) Calculus
B) Linear algebra
C) Probability
D) Statistics

A) Incorrect.
B) Incorrect.
C) Correct. Probability is a mathematical way of measuring uncertainty. Based on the data, a machine learning model will use measurements of uncertainty to make a decision.
D) Incorrect.

4 / 18
Machine learning can be used to sort unlabeled data into groups.

What is this known as?

A) Classification
B) Clustering
C) Grouping
D) Prediction

A) Incorrect.
B) Correct. Clustering can be used to group unlabeled data based on similar features, where there are no labels to easily sort or classify data into specific categories.
C) Incorrect.
D) Incorrect.

5 / 18
Which **two** of the following languages are commonly used in machine learning?

Please remember to choose 2 answers.

A) CSS
B) MATLAB
C) Python
D) SQL

A) Incorrect.
B) Correct. Python, MATLAB, Java and R are common languages used in machine learning. SQL is used for structuring data within an database. CSS and HTML5 are used to develop and style web content.
C) Correct. Python, MATLAB, Java and R are common languages used in machine learning. SQL is used for structuring data within an database. CSS and HTML5 are used to develop and style web content.
D) Incorrect.

6 / 18
What is required in order to train a machine learning model?

A) Neural network
B) Teacher
C) Test data
D) Visualization tools

A) Incorrect.
B) Incorrect.
C) Correct. In order to train a machine learning model, you need data and an algorithm.
D) Incorrect.

7 / 18
Which of the following **two** frameworks can be used to develop machine leaning models?

Please remember to choose 2 answers.

A) Google analytics
B) Minecraft
C) Scikit-Learn
D) TensorFlow

A) Incorrect.
B) Incorrect.
C) Correct. TensorFlow and Scikit-Learn are two open-source frameworks that can be used to build machine learning.
D) Correct. TensorFlow and Scikit-Learn are two open-source frameworks that can be used to build machine learning.

8 / 18
Satpal has been developing an application (App) that can be used to order food from different restaurants and have it delivered straight to your home. As part of the functionality, he has built in a machine learning model that uses regression to provide the user with an 'expected delivery time' for the food based on the time of day, the distance between the restaurant to the intended location, and the average delivery time.

The algorithm being used has been configured to compare the two variables 'time of day' and 'average delivery time' in order to make its prediction.

When testing the App, he has found that the predicted 'expected delivery time' seems incredibly long based on his location to the restaurant.

What is probably the issue?

A) The average delivery time
B) The average recorded speed of the driver
C) The choice of algorithm
D) The variables being compared in the data

A) Incorrect.
B) Incorrect.
C) Incorrect.
D) Correct. Satpal has not included the variable 'distance' in his model. To calculate the expected delivery time, the algorithm should compare the three variables 'time of day, 'distance' and 'average delivery time' to be able to make a more accurate prediction.

9 / 18
Roisin is training her machine learning model using unlabeled data and no training data.

What type of type of approach would she use?

A) Reinforcement learning
B) Semi-supervised learning
C) Supervised learning
D) Unsupervised learning

A) Incorrect.
B) Incorrect.
C) Incorrect.
D) Correct. Unsupervised learning is often used to derive insight from unlabeled data, and when there is no target outcome yet defined. A common algorithm used in unsupervised learning is clustering.

10 / 18
Which **two** of the following problems can be solved through classification?

Please remember to choose 2 answers.

A) Grouping sets of unlabeled data to identify different customer segments
B) Identifying an image based on specific features in the data
C) Making predictions on the number of cases of a virus in a particular area
D) Sorting emails into 'received' and 'spam'

A) Incorrect.
B) Correct. Classification can be used to sort labeled data such as e-mails into specific categories based on certain features. It can also be used in image recognition using a neural network where certain features are identified and compared to a stored image to classify what the image is.
C) Incorrect.
D) Correct. Classification can be used to sort labeled data such as e-mails into specific categories based on certain features. It can also be used in image recognition using a neural network where certain features are identified and compared to a stored image to classify what the image is.

11 / 18
Which of the following is open-source software developed by Google to design machine learning models?

A) Google analytics
B) Scikit-Learn
C) TensorFlow
D) Weka

A) Incorrect.
B) Incorrect.
C) Correct. TensorFlow is commonly used to design machine learning models.
D) Incorrect.

12 / 18
Which type of algorithm is commonly used within deep learning?

A) K-nearest neighbors
B) Linear regression
C) Naïve Bayes
D) Neural network

A) Incorrect.
B) Incorrect.
C) Incorrect.
D) Correct. Neural networks are commonly used within deep learning due to their multiple layers that allow for more complex problems to be solved.

13 / 18
What is supervised learning?

A) Training a machine learning model through reinforcement
B) Training a machine learning model using unlabeled data
C) Training a machine learning model where the desired output is known
D) Training a machine learning model where the output is unknown

A) Incorrect.
B) Incorrect.
C) Correct. In supervised learning we will often have labeled data that we can use to train our machine learning model, so that when it is provided with similar data it knows how to classify it. Therefore, the desired output is already known.
D) Incorrect.

14 / 18
Dale is wishing to develop an application of machine learning that is able to sort different types of user requests that have been manually input into a system. Due to the volume of user requests it is often time-consuming for an individual to read through each request. It is not always easy to quantify or prioritize requests based on how many times the same type of request is made.

Dale has observed that there are specific words that regularly feature in certain types of user requests that could be used to identify them. He would therefore like a machine learning model to read through each request and sort them into defined categories based on whether these specific words feature.

What type of approach should he use to solve the problem?

A) Classification
B) Clustering
C) Regression
D) Reinforcement

A) Correct. An algorithm such as Naive Bayes could be used to solve this problem, whereby the data could be classified based on certain characteristics i.e. the words contained in each user request.
B) Incorrect.
C) Incorrect.
D) Incorrect.

15 / 18
When would you use unsupervised learning?

A) When our algorithm does not require our input
B) When we are too busy to oversee the process
C) When we have sets of labeled data
D) When we have sets of unlabeled data

A) Incorrect.
B) Incorrect.
C) Incorrect.
D) Correct. Unsupervised learning can be used to train a machine learning model when we have sets of unlabeled data and we therefore do not know what the desired output is. A clustering algorithm is typically used within unsupervised learning.

16 / 18
Which of the following algorithms can be used within clustering?

A) K-means
B) K-nearest neighbors
C) Linear regression
D) Logistic regression

A) Correct. K relates to the number of clusters we wish to group our data into. By assigning data points within we can then group data based on how close they are to each data point, so clustering data.
B) Incorrect.
C) Incorrect.
D) Incorrect.

17 / 18
What is the purpose of data pre-processing?

Please remember to choose 2 answers.

A) To clean the data to ensure it is suitable for training a machine learning model
B) To identify features or target values in the data that can be used to create training data
C) To identify the types of data required to solve the problem
D) To present the data using a graph or chart

A) Correct. Data selection is the process of identifying the types of data required to solve our problem. Data visualization is later used to display the data in a graphical format to make it easier to analyze the data before training a model.
B) Correct. Data selection is the process of identifying the types of data required to solve our problem. Data visualization is later used to display the data in a graphical format to make it easier to analyze the data before training a model.
C) Incorrect.
D) Incorrect.

18 / 18
What is the correct order of steps in the machine learning process?

A) 1. Problem identification
2. Data pre-processing
3. Data selection
4. Testing
5. Training

B) 1. Problem identification
2. Data pre-processing
3. Data selection
4. Training
5. Testing

C) 1. Problem identification
2. Data selection
3. Data pre-processing
4. Training
5. Testing

D) 1. Problem identification
2. Training
3. Data selection
4. Data pre-processing
5. Testing

A) Incorrect.

B) Incorrect.

C) Correct. The final stages of this process would be to review the results before deploying your model. You would typically undertake data visualization prior to training in order to analyze your data in order to make better sense of it before using it to train your model.

D) Incorrect.

Evaluation

The table below shows the correct answers to the questions in this sample exam.

Question	Answer	Question	Answer
1	C	**10**	B & D
2	B	**11**	C
3	C	**12**	D
4	B	**13**	C
5	B & C	**14**	A[1]
6	C	**15**	D
7	C & D	**16**	A
8	D[2]	**17**	A & B
9	D	**18**	C

[1] Question is scored with two points if it is marked correct.
[2] Question is scored with two points if it is marked correct.

Contact EXIN

www.exin.com