Table of Contents

Points to Note

The rest of this document represents a sample test paper for AI/ML knowledge.

Instructions as seen by the Candidate

Machine Learning Knowledge

This is an optional test for AI and machine learning knowledge. Your performance in this test will be used to decide whether you will be interviewed by the the AI/ML team at Icertis

You will be asked 15 objective questions questions related to machine learning concepts, knowledge of algorithms, and the underlying mathematical concepts. You have 20 minutes to answer them.

Click on "Start this activity" (below) to get started.

Section 1: Machine Learning Algorithms.

This section has 5 questions

--

Train vs Test #1:

A neural network for an application produces 90% accuracy on the training data, but only 60% accuracy on the test data. Which of the following changes to the neural net is most likely to improve the accuracy on the test data. (Assuming the entire network is trained again after the change.)

Choose the option closest to the correct answer (only one):

1. Increasing the size of the training data
2. Increasing the number of layer in the neural network
3. Increasing the number of neurons in the hidden layer
4. Increasing the size of the test data
5. Increasing the number of iterations/epochs

--

Train vs Test #2:

A neural network for an application produces 60% accuracy on the training data, and around 55% accuracy on the test data. Which of the following changes to the neural net is most likely to improve the accuracy for both the training and test data. (Assuming the entire network is trained again after the change.)

Choose the option closest to the correct answer (only one):

1. Increasing the size of the training data
2. Changing the loss function
3. Increasing the number of neurons in the neural network
4. Increasing the size of the test data
5. Increasing the number of iterations/epochs

--

Gradient Descent:

In the context of neural network based machine learning algorithms, where is gradient descent primarily used?

Choose the option closest to the correct answer (only one):

1. It is the technique used to decide how to update the weights/hyperparameters after a batch of test data has been processed
2. It is a method used to decide how many iterations/epochs to run the training for
3. It is the technique used to decide the error/accuracy of the current model
4. It is the function in a neural network applied to the weighted inputs to produce the output
5. None of the above

--

Feature Engineering:

In the context of machine learning, extracting and preparing the features of the input data is an important part of the process. Which of the following techniques represent techniques that are commonly used as a part of feature engineering?

Choose one or more options:

1. Splitting a data item into multiple different features, for example converting a date column into 3 different columns: year, month, days etc.
2. Dropping or capping outliers in the data
3. Converting numerical data to string data for easier concatenation
4. Scaling and normalizing numerical data so that all values range from 0 to 1.
5. None of the above

--

Learning Rate:

In the context of neural network based machine learning algorithms, which of the following statements about the learning rate are true?

Choose one or more options:

1. It should be as low as possible
2. If it is too low the model will take too long to converge
3. It should be low in the beginning and increase as the model converges
4. It should be as high as possible
5. If it is too high the model might jump over minima and miss them

Section 2: ML and Data Science.

This section has 5 questions

--

Decision Tree vs Random Forest:

Which of the following represent advantages of decision tree learning (DTL) algorithms over the use of random forests for a classification task

Choose one or more options:

1. A typical DTL algorithm will have higher variance compared to random forests
2. DTL requires labeled training data while random forests do not
3. In case a human requires an explanation of why a certain input produced a certain output, the explanation would be easier/more understandable in case of DTL
4. A typical DTL algorithm will have have higher bias compared to random forests
5. None of the above: random forests are superior to DTL in all ways

--

Principal components:

Consider matrix X which consists of n rows with p columns each. Each row represents a different customer, and each column represents a different feature of that customer. (Assume that the values have already been adjusted so as to be directly usable for a PCA transformation.) Assume that $T = X W$ represents the principal components decomposition of X where W is the matrix which has p columns representing the eigenvectors of X^TX. Which of the following statements are true?

Choose one or more options:

1. The result T will be a $p \times p$ matrix
2. Taking the first k rows of T will result in a dimentionality reduction of the original data form n to k
3. The result T will be a $n \times n$ matrix
4. The result T will consist of n rows of p columns each
5. Taking the first k columns of T will result in a dimentionality reduction of the original data form p to k

--

Supervised vs Unsupervised Learning:

Which of the following statements about supervised and unsupervised learning are true?

Choose one or more options:

1. In supervised learning, the training data set contains both inputs as well as desired outputs
2. The k-means algorithm is an example of supervised learning
3. Clustering is an example of unsupervised learning
4. Supervised learning uses statistical techniques, while unsupervised learning uses neural networks
5. In supervised learning, a human is given an chance to modify the final output, whereas unsupervised learning is fully automated

--

k-nearest neighbors:

A data scientist is trying to use a k-NN algorithm to categorize the customers of the bank into 10 different categories (based on value of the customer to the bank, in terms of revenue). For each customer 20 different features have already been extracted, and the k-NN is run (with k=10) on these features. Unfortunately, the k-NN algorithm is not giving good results (the categorization isn't very good). Which of the following represents a reasonable approach to fixing the problem?

Choose one or more options:

1. Start with `k=20`, and then run the entire algorithm repeatedly decreasing `k` by 1 each time
2. Decrease the number of dimensions by using PCA or another dimensionality reduction technique
3. Start with `k=2`, and then run the entire algorithm repeatedly increasing `k` by 1 each time
4. Increase the number of categories to 20
5. Use neural networks, because k-NN is not an effective algorithm for this kind of problems

--

Decision Tree vs Neural Networks:

Which of the following represent advantages of decision tree learning (DTL) algorithms over the use of neural networks (NN) for a classification task

Choose one or more options:

1. A typical DTL implementation would require fewer computational resources than NN
2. DTL can handle numerical or categorical values, whereas categories values need to be converted to numbers in case of NN
3. In case a human requires an explanation of why a certain input produced a certain output, the explanation would be easier/more understandable in case of DTL
4. DTL requires labeled training data while NN doesn't
5. None of the above: modern NN are superior to DTL in all ways

Section 3: Maths for ML.

This section has 5 questions

--

Eigenvectors:

Consider the following:

Let $A = \begin{bmatrix} 4 & -2 & 1 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix}$, $v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $v_3 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ and $v_4 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$

Which of the vectors are eigenvectors of A?

Choose one or more options:

1. v2
2. v4
3. v3
4. v1
5. None of the above

--

Eigenvalue:

Consider the following:

$\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ is an eigenvector of $\begin{bmatrix} 4 & -2 & 1 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix}$

What is the corresponding eigenvalue?

Choose the option closest to the correct answer (only one):

1. 2
2. 4
3. $[2\ 4\ 4]^T$
4. [1 2 2]
5. None of the above

--

Maxima and Minima:

Consider the function $f(x) = 2x^3 - 15x^2 + 36x + 3$. Which of the following values of x represent maxima or minima for f(x)?

Choose one or more options:

1. 1
2. 3
3. 2
4. 0

5. 15

--

Even and Odd:

Let `n` be a positive integer, such that `(n+3)(n+5)` is odd. Let `m = (n+4)(n+6)`. Which of the following statements *must* be true?

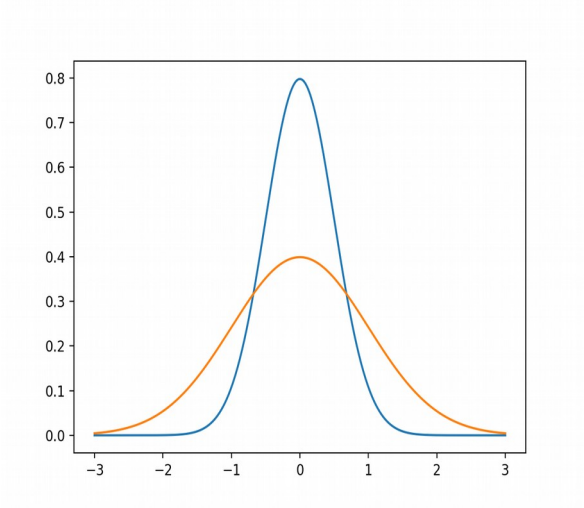
Choose one or more options:

1. `n+6` is divisible by 2
2. `m` is divisible my 8
3. Exactly one of `n+4` and `n+6` must be divisible by 4
4. `n+4` is divisible by 2
5. None of the above

--

Variance:

Consider the following graph:



Both the curves in the image have a normal distribution with the same mean (0). Which of the curves has a higher variance?

Choose the option closest to the correct answer (only one):

1. Blue
2. There is insufficient information in the image
3. Orange
4. Both curves have the same variance
5. The question is irrelevant because variance for normal distribution is undefined

Answer Key

Section 1: Machine Learning Algorithms

1. 1
2. 5
3. 1
4. 1, 2, 4
5. 2, 5

Section 2: ML and Data Science

1. 1, 3
2. 4, 5
3. 1, 3
4. 2
5. 1, 2, 3

Section 3: Maths for ML

1. 1, 4
2. 1
3. 2, 3
4. 1, 2, 3, 4
5. 3