Top 40 Machine Learning Questions & Answers for Beginners and Experts (Updated 2025)

Ankit Gupta
Last Updated : 04 Apr, 2025

10

Introduction

Machine Learning, a prominent part of Artificial Intelligence, is currently one of the most sought-after skills in data science. If you are a data scientist, you need to be good at python, SQL, and machine learning – no two ways about it. As part of DataFest 2017, we organized various skill tests so that data scientists could assess themselves on these critical skills. These tests included Machine Learning, Deep Learning, **Time Series problems**, and **Probability**. This article will lay out the solutions to the **machine learning Questions and Answers** to skill test and other important data science interview questions.In this article you will get to know and machine learning exam questions and answers and how they are impacting.

Table of contents

About the Skill Test

More than 1350 people registered for the Machine Learning skill test. The test was designed to test your conceptual knowledge in data science and machine learning. This article gives you a chance to test yourself in case you missed the real-time test. Here are the **leaderboard** rankings for all the participants in the Machine Learning Skilltest. You can also check out our **online training in machine learning**.

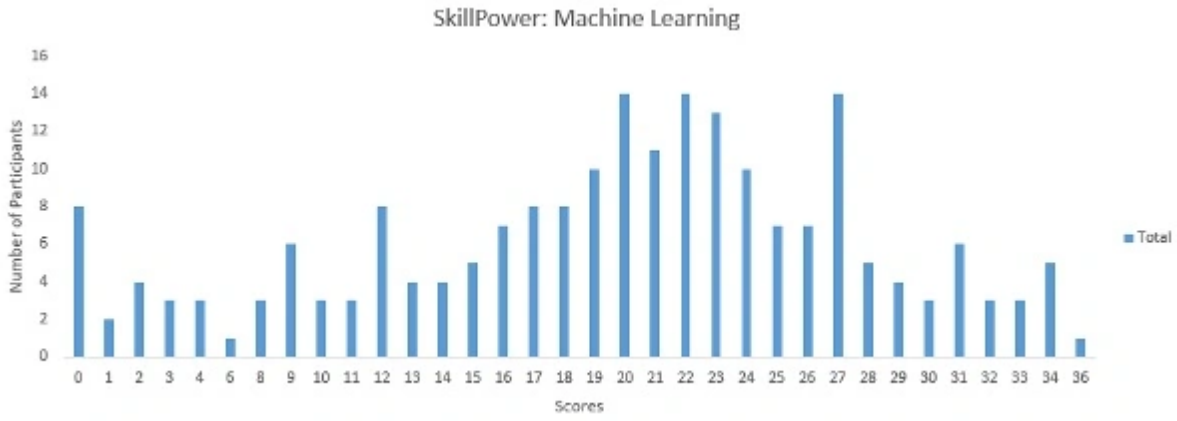
more. It also quizzes you on **Statistics** concepts such as Normal distribution, p-value, etc.

These data science questions, along with hundreds of others, are part of our '**Ace Data Science Interviews**' course. It's a comprehensive guide, with tons of resources, to crack data science interviews and land your dream job! And if you're someone who's just starting out their data science journey, then do check out our most comprehensive program to master Machine Learning:

- **Certified AI & ML BlackBelt+ Program**
- **Machine Learning Certification Course for Beginners**

Take the test and find where you stand!

Below is the distribution of the overall scores that will help you evaluate your performance.



You can access the **final scores** here. More than 210 people participated in the machine learning skill test, and the highest score obtained was 36. Here are a

Machine Learning Questions & Answers

Q1) A feature F1 can take a certain value: A, B, C, D, E, or F, which represents the grades of students from a college.

Which of the following statement is true in the following case?

A) Feature F1 is an example of a nominal variable.

B) Feature F1 is an example of an ordinal variable.

C) It doesn't belong to any of the above categories.

D) Both of these

Solution: (B)

Ordinal variables are the variables that have some order in their categories. For example, grade A should be considered a high grade than grade B.

Q2) Which of the following is an example of a deterministic algorithm?

A) PCA

B) K-Means clustering

C) KNN

D) None of the above

Solution: (A)

A deterministic algorithm is one in which output does not change on different runs. PCA would give the same result if we run again, but not k-means clustering

A) TRUE

B) FALSE

Solution: (A)

$Y=X2$. Note that they are not only associated, but one is a function of the other, and the Pearson correlation between them is 0.

Q4) Which of the following statement(s) is / are true for Gradient Descent (GD) and Stochastic Gradient Descent (SGD)?

1. **In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.**
2. **In SGD, you must run through all the samples in your training set for a single parameter update in each iteration.**
3. **In GD, you either use the entire data points or a subset of training data to update a parameter in each iteration.**

A) Only 1

B) Only 2

C) Only 3

D) 1 and 2

E) 2 and 3

F) 1, 2 and 3

Solution: (A)

Q5) Which of the following hyper-parameter (s), when increased, may cause the random forest to over fit the data?

1. **Number of Trees**
2. **Depth of Tree**
3. **Learning Rate**

A) Only 1

B) Only 2

C) Only 3

D) 1 and 2

E) 2 and 3

F) 1, 2 and 3

Solution: (B)

Usually, if we increase the depth of the tree, it will cause overfitting. The learning rate is not a hyperparameter in a random forest. An increase in the number of trees will cause under fitting.

Q6) Suppose you want to develop a machine learning algorithm that predicts the number of views on the articles in a blog.

Your data analysis is based on features like author name, number of articles written by the same author, etc. Which of the following evaluation metrics would you choose in that case?

A) Only 1

B) Only 2

C) Only 3

D) 1 and 3

E) 2 and 3

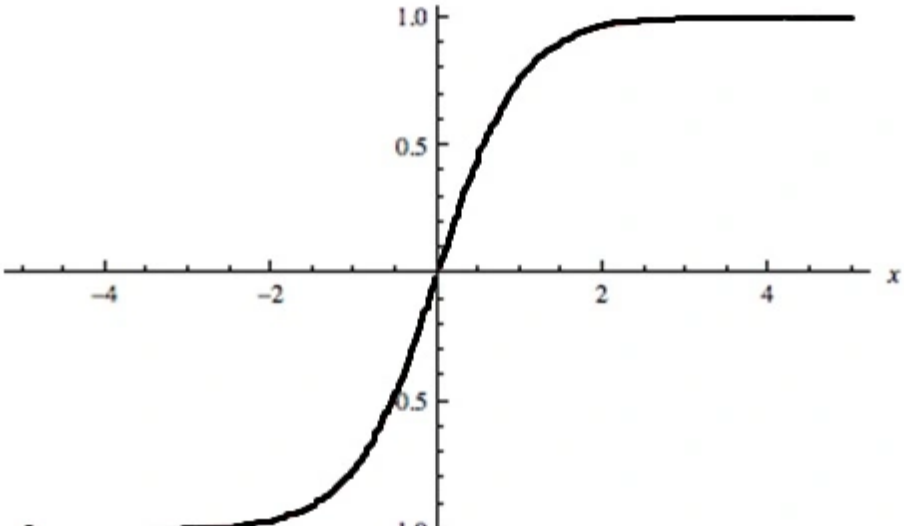
F) 1 and 2

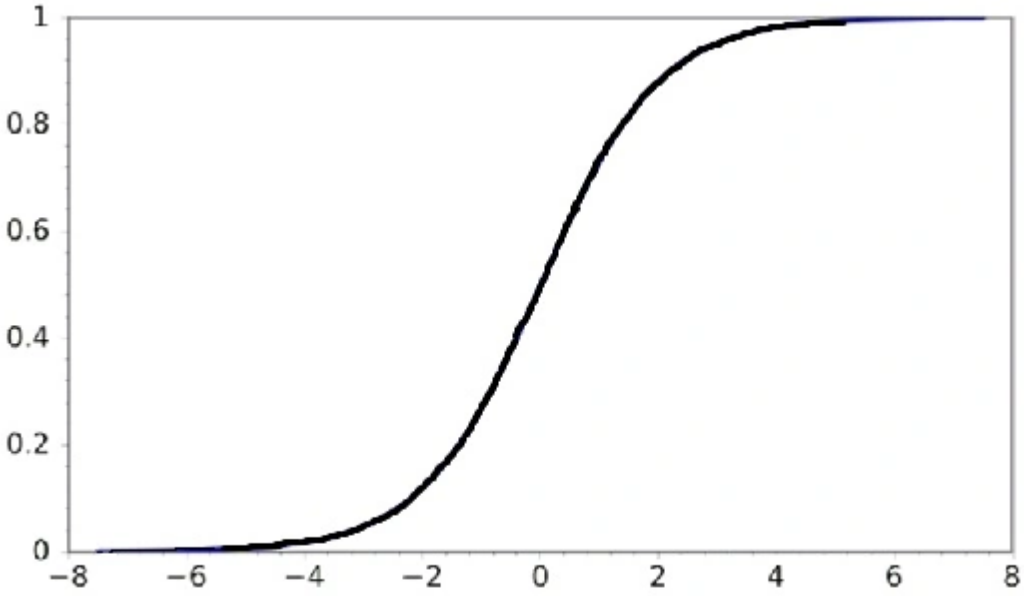
Solution:(A)

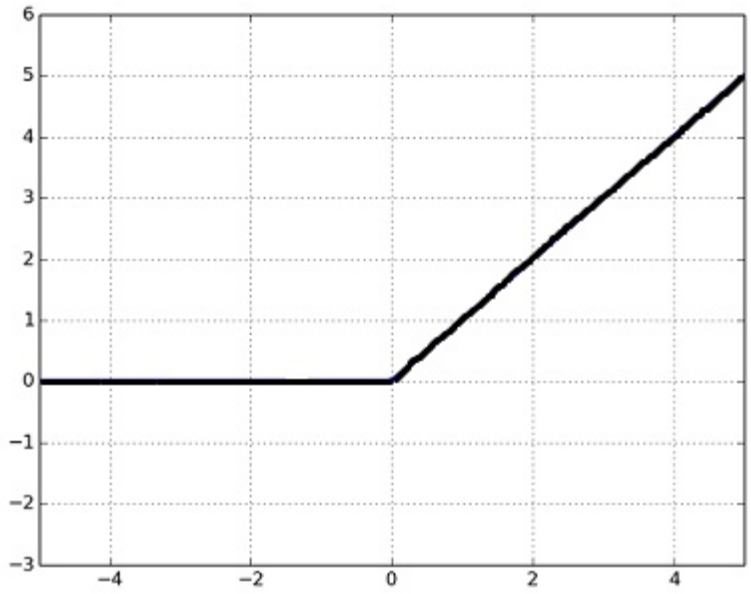
You can think that the number of views of articles is the continuous target variable that falls under the regression problem. So, the mean of squared error will be used as an evaluation metric.

Q7) Given below are three images (1,2,3). Which of the following option is correct for these images?

A)







A) 1 is tanh, 2 is ReLU, and 3 is SIGMOID activation functions.

B) 1 is SIGMOID, 2 is ReLU, and 3 is tanh activation functions.

C) 1 is ReLU, 2 is tanh, and 3 is SIGMOID activation functions.

D) 1 is tanh, 2 is SIGMOID, and 3 is ReLU activation functions.

The RELU function has the range of [0, infinity].

So Option D is the right answer.

Q8) Below are the 8 actual target variable values in the training file.

[0,0,0,1,1,1,1,1]

What is the entropy of the target variable?

A) -(5/8 log(5/8) + 3/8 log(3/8))

B) 5/8 log(5/8) + 3/8 log(3/8)

C) 3/8 log(5/8) + 5/8 log(3/8)

D) 5/8 log(3/8) – 3/8 log(5/8)

Solution: (A)

The formula for entropy is So the answer is A.

Q9) Let's say you are working with categorical feature(s), and you have not looked at the distribution of the categorical variable in the test data.

You want to apply one hot encoding (OHE) on the categorical feature(s). What challenges may you face if you have applied OHE on a categorical variable of the training dataset?

A) All categories of the categorical variables are not present in the test dataset.

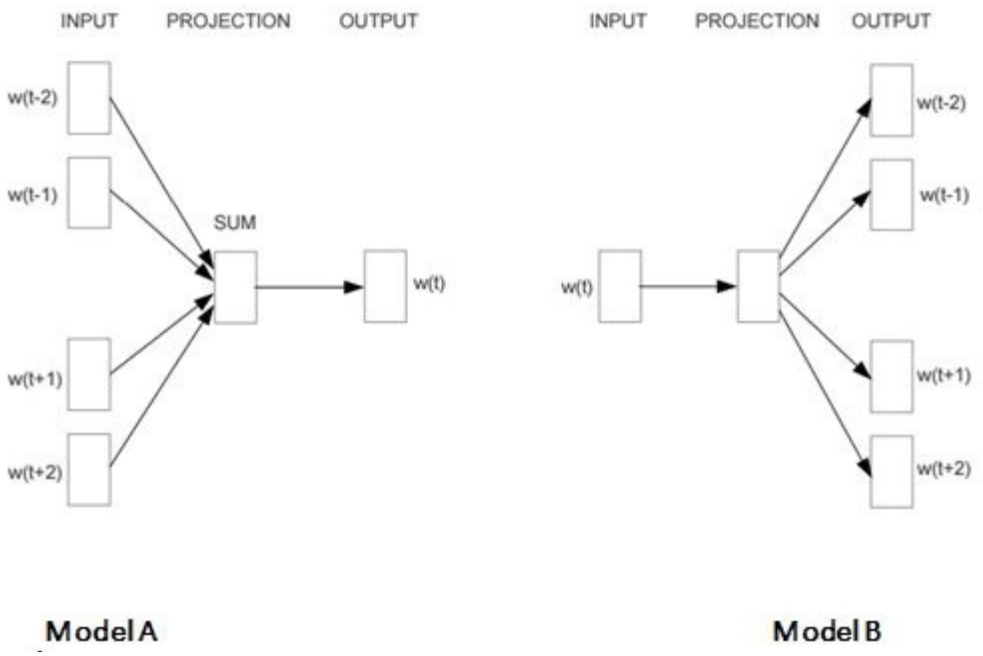
B) The frequency distribution of categories is different in the train compared to the test dataset.

Solution: (D)

Both are true. The OHE will fail to encode the categories which is present in the test but not in the train, so it could be one of the main challenges while applying OHE. The challenge given in option B is also true. You need to be more careful while applying OHE if frequency distribution isn't the same in the train and the test.

Q10) Skip-gram model is one of the best models used in the Word2vec algorithm for embedding words.

Which one of the following models depicts the skip-gram model?



A) A

B) B

Both models (model1 and model2) are used in the Word2vec algorithm. Model1 represents a CBOW model, whereas Model2 represents the Skip-gram model.

Q11) Let's say you are using activation function X in hidden layers of a neural network.

At a particular neuron for any given input, you get the output as "-0.0001".

Which of the following activation function could X represent?

A) ReLU

B) tanh

C) SIGMOID

D) None of these

Solution: (B)

The function is a tanh because this function output range is between (-1,-1).

Q12) [True or False] LogLoss evaluation metric can have negative values.

A) TRUE

B) FALSE

Solution: (B)

Log loss cannot have negative values.

1. **Type I error is known as a false positive, and Type II error is known as a false negative.**
2. **Type I error is known as a false negative, and Type II error is known as a false positive.**
3. **Type I error occurs when we reject a null hypothesis when it is actually true.**

A) Only 1

B) Only 2

C) Only 3

D) 1 and 2

E) 1 and 3

F) 2 and 3

Solution: (E)

In statistical hypothesis testing, a type I error is the incorrect rejection of a true null hypothesis (a “false positive”), while a type II error is incorrectly retaining a false null hypothesis (a “false negative”).

Q14) Which of the following is/are one of the important step(s) to pre-process the text in NLP-based projects?

1. **Stemming**
2. **Stop word removal**

C) 2 and 3

D) 1,2 and 3

Solution: (D)

Stemming is a rudimentary rule-based process of stripping the suffixes (“ing”, “ly”, “es”, “s”, etc.) from a word.

Stop words are those words which will have not relevant to the context of the data, for example, is/am/are.

Object Standardization is also a good way to pre-process the text.

Q15) Suppose you want to project high-dimensional data into lower dimensions.

The two most famous dimensionality reduction algorithms used here are PCA and t-SNE. Let’s say you have applied both algorithms respectively on data “X”, and you got the datasets “X_projected_PCA” , “X_projected_tSNE”.
Which of the following statements is true for “X_projected_PCA” & “X_projected_tSNE”?

A) X_projected_PCA will have interpretation in the nearest neighbor space.

B) X_projected_tSNE will have interpretation in the nearest neighbor space.

C) Both will have interpretation in the nearest neighbor space.

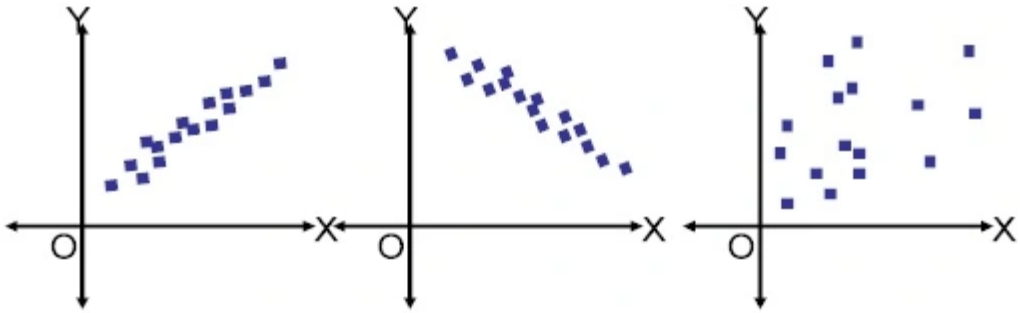
D) None of them will have interpretation in the nearest neighbor space.

Solution: (B)

have interpretation in the nearest neighbor space. But in the case of PCA, it is not the case.

Context for Questions 16-17

Given below are three scatter plots for two features (Images 1, 2 & 3, from left to right).



Q16) In the above images, which of the following is/are examples of multi-collinear features?

A) Features in Image 1

B) Features in Image 2

C) Features in Image 3

D) Features in Images 1 & 2

E) Features in Images 2 & 3

F) Features in Images 3 & 1

Solution: (D)

Q17) In the previous question, suppose you have identified multi-collinear features. Which of the following action(s) would you perform next?

1. **Remove both collinear variables.**
2. **Instead of removing both variables, we can remove only one variable.**
3. **Removing correlated variables might lead to a loss of information. In order to retain those variables, we can use penalized regression models like ridge or lasso regression.**

A) Only 1

B) Only 2

C) Only 3

D) Either 1 or 3

E) Either 2 or 3

Solution: (E)

You cannot remove both features because after removing them, you will lose all of the information. So you should either remove only 1 feature or use a **regularization** algorithm like L1 and L2.

Q18) Adding a non-important feature to a linear regression model may result in.

1. **Increase in R-square**

C) Either 1 or 2

D) None of these

Solution: (A)

After adding a feature in the feature space, whether that feature is an important or unimportant one, the R-squared always increases.

Q19) Suppose you are given three variables X, Y, and Z. The Pearson correlation coefficients for (X, Y), (Y, Z), and (X, Z) are C1, C2 & C3, respectively.

Now, you have added 2 in all values of X (i.e., new values become X+2), subtracting 2 from all values of Y (i.e., new values are Y-2), and Z remains the same. The new coefficients for (X,Y), (Y,Z), and (X,Z) are given by D1, D2 & D3, respectively. How do the values of D1, D2 & D3 relate to C1, C2 & C3?

A) D1= C1, D2 < C2, D3 > C3

B) D1 = C1, D2 > C2, D3 > C3

C) D1 = C1, D2 > C2, D3 < C3

D) D1 = C1, D2 < C2, D3 < C3

E) D1 = C1, D2 = C2, D3 = C3

F) Cannot be determined

Solution: (E)

Correlation between the features won't change if you add or subtract a value in

The majority class is observed 99% of the time in the training data. Your model has 99% accuracy after taking the predictions on the test set. Which of the following is true in such a case?

1. **The accuracy metric is not a good idea for imbalanced class problems.**
2. **The accuracy metric is a good idea for imbalanced class problems.**
3. **Precision and recall metrics are good for imbalanced class problems.**
4. **Precision and recall metrics aren't good for imbalanced class problems.**

A) 1 and 3

B) 1 and 4

C) 2 and 3

D) 2 and 4

Solution: (A)

Refer the question number 4 from this **article**.

Q21) In ensemble learning (i.e., bagging or boosting), you aggregate the predictions for weak learners so that an ensemble of these models will give a better prediction than the prediction of individual machine learning models.

Which of the following statements is / are true for weak learners used in the ensemble model?

1. **They don't usually overfit.**

A) 1 and 2

B) 1 and 3

C) 2 and 3

D) Only 1

E) Only 2

F) None of the above

Solution: (A)

Weak learners are sure about a particular part of a problem. So, they usually don’t overfit, which means that weak learners have low variance and high bias.

Q22) Which of the following options is/are true for K-fold cross-validation?

1. **An increase in K will result in a higher time required to cross-validate the result.**
2. **Higher values of K will result in higher confidence in the cross-validation result as compared to a lower value of K.**
3. **If K=N, then it is called Leave one out cross validation, where N is the number of observations.**

A) 1 and 2

B) 2 and 3

C) 1 and 3

A larger k-value means less bias towards overestimating the truly expected error (as training folds will be closer to the total dataset) and higher running time (as you are getting closer to the limit case: Leave-One-Out CV). We also need to consider the variance between the k folds accuracy while selecting the k.

Context for Questions 23-24

Cross-validation is an important step in machine learning for hyper parameter tuning. Let's say you are tuning a hyper-parameter "max_depth" for GBM by selecting it from 10 different depth values (values are greater than 2) for a tree-based model using 5-fold cross-validation.
Time taken by an algorithm for training (on a model with max_depth 2) 4-fold is 10 seconds, and for the prediction on remaining 1-fold is 2 seconds.
Note: Ignore hardware dependencies from the equation.

Q23) Which of the following option is true for overall execution time for 5-fold cross-validation with 10 different values of "max_depth"?

A) Less than 100 seconds

B) 100 – 300 seconds

C) 300 – 600 seconds

D) More than or equal to 600 seconds

E) None of the above

F) Can't estimate

Solution: (D)

seconds. But training and testing a model on a depth greater than 2 will take more time than depth “2”, so overall timing would be greater than 600.

Q24) In the previous question, if you train the same algorithm for tuning 2 hyperparameters, say “max_depth” and “learning_rate”.

You want to select the right value against “max_depth” (from the given 10 depth values) and learning rate (from the given 5 different learning rates). In such cases, which of the following will represent the overall time?

A) 1000-1500 second

B) 1500-3000 Second

C) More than or equal to 3000 Second

D) None of these

Solution: (D)

Explanation is the same as above.

Q25) Given below is a scenario for training error TE and Validation error VE for a machine learning algorithm M1.

You want to choose a hyperparameter (H) based on TE and VE.

H	TE	VE
1	105	90
2	200	85

Which value of H will you choose based on the above table?

A) 1

B) 2

C) 3

D) 4

E) 5

Solution: (D)

Looking at the table, option D seems the best

Q26) What would you do in PCA to get the same projection as SVD?

A) Transform data to zero mean

B) Transform data to zero median

C) Not possible

D) None of these

Solution: (A)

When the data has a zero mean vector, PCA will have the same projections as SVD; otherwise, you have to center the data first before taking SVD.

Context for Questions 27-28

Assume there is a black box algorithm that takes training data with multiple observations (t1, t2, t3 ..... tn) and a new observation (q1)

You can also think that this black box algorithm is the same as 1-NN (1-nearest neighbor).

Q27) Is it possible to construct a k-NN classification algorithm based on this black box alone?

Note: Where n (number of training observations) is very large compared to k.

A) TRUE

B) FALSE

Solution: (A)

In the first step, you pass an observation (q1) in the black box algorithm, so this algorithm returns the nearest observation and its class.
In the second step, you go through the nearest observation from train data and again input the observation (q1). The black box algorithm will again return the nearest observation and its class.
You need to repeat this procedure k times.

Q28) Instead of using a 1-NN black box, we want to use the j-NN (j>1) algorithm as a black box. Which of the following option is correct for finding k-NN using j-NN?

1. **J must be a proper factor of k**
2. **J > k**

Solution: (A)

Explanation is the same as above.

Q29) Suppose you are given 7 Scatter plots 1-7 (left to right), and you want to compare Pearson correlation coefficients between variables of each scatterplot.

Which of the following is in the right order?

1. **1<2<3<4**
2. **1>2>3 > 4**
3. **7<6<5<4**
4. **7>6>5>4**

Q30) Which of the following option is / are true for the interpretation of log loss as an evaluation metric?

$$logLoss = \frac{-1}{N}\sum_{i=1}^{N}(y_i(logp_i) + (1-y_i)log(1-p_i))$$

1. **If a classifier is confident about an incorrect classification, then log-loss will penalize it heavily.**

A) 1 and 3

B) 2 and 3

C) 1 and 2

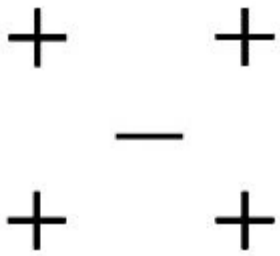
D) 1,2 and 3

Solution: (D)

Options are self-explanatory.

Context for Questions 31-32

Below are five samples given in the dataset.



Note: Visual distance between the points in the image represents the actual distance.

Q31) Which of the following is leave-one-out cross-validation accuracy for 3-NN (3-nearest neighbor)?

A) 0

B) 0.4

C) 0.8

In Leave-One-Out cross-validation, we will select (n-1) observations for training and 1 observation of validation. Consider each point as a cross-validation point and then find the 3 nearest points to this point. So if you repeat this procedure for all points, you will get the correct classification for all positive classes given in the above figure, but the negative classes will be misclassified. Hence you will get 80% accuracy.

Q32) Which of the following value of K will have the least leave-one-out cross-validation accuracy?

A) 1NN

B) 3NN

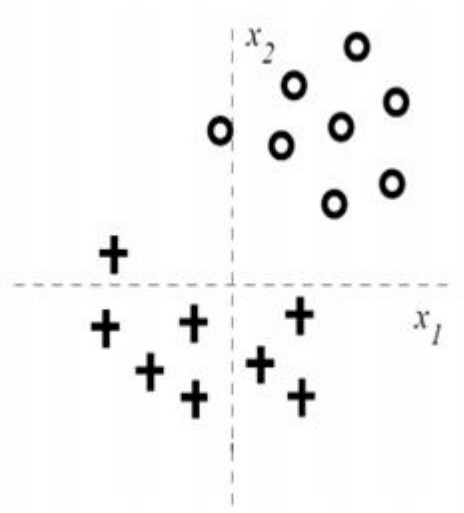
C) 4NN

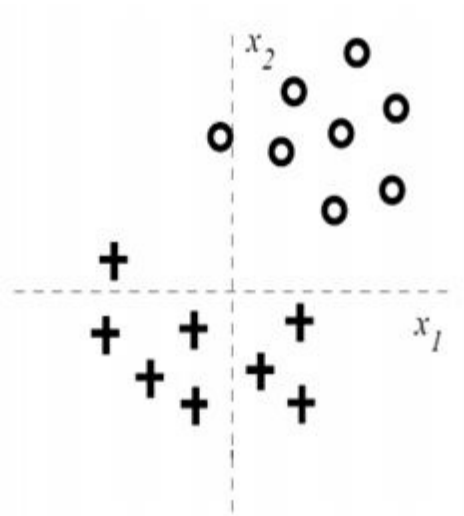
D) All have the same leave-one-out error

Solution: (A)

Each point will always be misclassified in 1-NN which means that you will get 0% accuracy.

Q33) Suppose you are given the below data, and you want to apply a logistic regression model for classifying it into two given classes.





You are using logistic regression with L1 regularization.

$$\sum_{i=1}^{n} \log P(y_i|x_i, w_0, w_1, w_2) - C(|w_1| + |w_2|).$$

Where C is the regularization parameter, and w1 & w2 are the coefficients of x1 and x2.

A) First, w2 becomes zero, and then w1 becomes zero

B) First, w1 becomes zero, and then w2 becomes zero

C) Both become zero at the same time

D) Both cannot be zero even after a very large value of C

Solution: (B)

By looking at the image, we see that even by just using x2, we can efficiently perform classification. So at first, w1 will become 0. As the regularization parameter increases more, w2 will come closer and closer to 0.

Q34) Suppose we have a dataset that can be trained with 100% accuracy with the help of a decision tree of depth 6.

Now consider the points below and choose the option based on these points.

Note: All other hyper parameters are the same, and other factors are not affected.

1. **Depth 4 will have high bias and low variance**
2. **Depth 4 will have low bias and low variance**

A) Only 1

B) Only 2

C) Both 1 and 2

D) None of the above

Solution: (A)

Q35) Which of the following options can be used to get global minima in k-Means Algorithm?

1. **Try to run an algorithm for different centroid initialization**
2. **Adjust the number of iterations**
3. **Find out the optimal number of clusters**

A) 2 and 3

B) 1 and 3

C) 1 and 2

D) All of above

Solution: (D)

All of the options can be tuned to find the global minima.

Q36) Imagine you are working on a project which is a binary classification problem.

You trained a model on the training dataset and got the below confusion matrix on the validation dataset.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual:		

1. **Accuracy is ~0.91**
2. **Misclassification rate is ~ 0.91**
3. **True Negative rate is ~0.95**
4. **True positive rate is ~0.95**

A) 1 and 3

B) 2 and 4

C) 1 and 4

D) 2 and 3

Solution: (C)

The Accuracy (correct classification) is (50+100)/165 which is nearly equal to 0.91.

The true Positive Rate is how many times you are predicting positive class correctly, so the true positive rate would be 100/105 = 0.95, also known as "Sensitivity" or "Recall"

Q37) For which of the following hyperparameters higher value is better for the decision tree algorithm?

1. **Number of samples used for split**
2. **Depth of tree**
3. **Samples for leaf**

D) 1, 2 and 3

E) Can't say

Solution: (E)

For all three options, A, B, and C, it is not necessary that if you increase the value of the parameter, the performance may increase. For example, if we have a very high value of depth of the tree, the resulting tree may overfit the data and would not generalize well. On the other hand, if we have a very low value, the tree may underfit the data. So, we can't say for sure that "higher is better."

Context for Questions 38-39

Imagine you have a 28 * 28 image, and you run a 3 * 3 convolution **neural network** on it with an input depth of 3 and an output depth of 8.
Note: Stride is 1, and you are using the same padding.

Q38) What is the dimension of the output feature map when you are using the given parameters?

A) 28 width, 28 height, and 8 depth

B) 13 width, 13 height, and 8 depth

C) 28 width, 13 height, and 8 depth

D) 13 width, 28 height, and 8 depth

Solution: (A)

Q39) What are the dimensions of the output feature map when you are using the following parameters?

A) 28 width, 28 height, and 8 depth

B) 13 width, 13 height, and 8 depth

C) 28 width, 13 height, and 8 depth

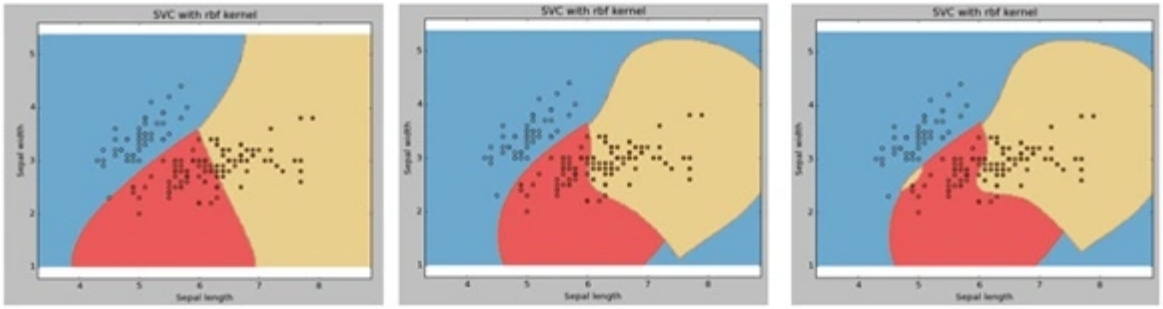
D) 13 width, 28 height, and 8 depth

Solution: (B)

Explanation is the same as above.

Q40) Suppose we were plotting the visualization for different values of C (Penalty parameter) in the SVM algorithm.

Due to some reason, we forgot to tag the C values with visualizations. In that case, which of the following option best explains the C values for the images below (1,2,3 left to right, so C values are C1 for image1, C2 for image2, and C3 for image3) in the case of an rbf kernel?



A) C1 = C2 = C3

Solution: (C)

Penalty parameter C of the error term. It also controls the trade-off between smooth decision boundaries and classifying the training points correctly. For large values of C, the optimization will choose a smaller-margin hyperplane. **Read here**.

Conclusion

I hope these questions and answers helped you test your knowledge and maybe learn a thing or two about Python, machine learning, and deep learning. You can find all the information about our upcoming skill tests and other **events** here. Machine learning **questions**.

Hope you like this article and getting understanding about the machine learning questions and machine learning exam questions and answers.You Will get a clear understanding about these questions and answers and it will make a better for learning the machine learning exam question and answers will help you to clear the interviews.

Key Takeways

- Concepts such as Supervised and Unsupervised learning, Neural Networks are very important to learn if you are aiming for a data scientist job.
- You must be good with data analysis skills, such as handling missing values and outliers

Other Useful Resources on Machine Learning

Here are some more articles and tutorials if you wish to explore machine learning further.

Machine Learning basics for a newbie

Deep Learning vs. Machine Learning – the essential differences you need to know!

Applied Machine Learning Course

Introduction to Data Science Course

Ace Data Science Interviews Course

Ankit Gupta

Ankit is currently working as a data scientist at UBS who has solved complex data mining problems in many domains. He is eager to learn more about data science and machine learning algorithms.

Intermediate | Interview Prep | Interviews | Machine Learning | Skilltest

Free Courses

Generative AI - A Way of Life

Explore Generative AI for beginners: create text and images, use top AI tools, learn practical skills, and ethics.

Getting Started with Large Language Models

Master Large Language Models (LLMs) with this course, offering clear guidance in NLP and model training made simple.

Building LLM Applications using Prompt Engineering

This free course guides you on building LLM apps, mastering prompt engineering, and developing chatbots with enterprise data.

Improving Real World RAG Systems: Key Challenges & Practical Solutions

Explore practical solutions, advanced retrieval strategies, and agentic RAG systems to improve

Microsoft Excel: Formulas & Functions

Master MS Excel for data analysis with key formulas, functions, and LookUp tools in this comprehensive course.

RECOMMENDED ARTICLES

45 Questions to Test a Data Scientist on Basics...

Top 100 Data Science Interview Questions and An...

Solutions for Skilltest Machine Learning : Reve...

30+ Most Important Data Science Interview Quest...

Top 20 Most Asked Machine Learning Interview Qu...

Top 15 Important Machine Learning Interview Que...

20 Data Science Interview Questions for a Beginner

Top 40 Machine Learning Interview Questions &am...

30 Questions to Test a Data Scientist on Tree B...

Responses From Readers

What are your thoughts?...

Submit reply

Amit Srivastava

For question 25, wouldnt Occam's Razor suggest choosing option 2. Its giving the same VE, but with a

Ankit Gupta

Hi Amit, It is true what you are saying but here hyperparameter H doesn't have any interpretation. So in such case you should choose the one which has lower training and validation error and also the close match. Best! Ankit Gupta

Free Courses

Generative AI | DeepSeek | OpenAI Agent SDK | LLM Applications using Prompt Engineering | DeepSeek from Scratch | Stability.AI | SSM & MAMBA | RAG Systems using LlamaIndex | Building LLMs for Code | Python | Microsoft Excel | Machine Learning | Deep Learning | Mastering Multimodal RAG | Introduction to Transformer Model | Bagging & Boosting | Loan Prediction | Time Series Forecasting | Tableau | Business Analytics | Vibe Coding in Windsurf | Model Deployment using FastAPI | Building Data Analyst AI Agent | Getting started with OpenAI o3-mini | Introduction to Transformers and Attention Mechanisms

Popular Categories

AI Agents | Generative AI | Prompt Engineering | Generative AI Application | News | Technical Guides | AI Tools | Interview Preparation | Research Papers | Success Stories | Quiz | Use Cases | Listicles

Generative AI Tools and Techniques

GANs | VAEs | Transformers | StyleGAN | Pix2Pix | Autoencoders | GPT | BERT | Word2Vec | LSTM | Attention Mechanisms | Diffusion Models | LLMs | SLMs | Encoder Decoder Models | Prompt Engineering | LangChain | LlamaIndex | RAG | Fine-tuning | LangChain AI Agent | Multimodal Models | RNNs | DCGAN | ProGAN | Text-to-Image Models | DDPM | Document Question Answering | Imagen | T5 (Text-to-Text Transfer Transformer) | Seq2seq Models | WaveNet | Attention Is All You Need (Transformer Architecture) | WindSurf | Cursor

Popular GenAI Models

Llama 4 | Llama 3.1 | GPT 4.5 | GPT 4.1 | GPT 4o | o3-mini | Sora | DeepSeek R1 | DeepSeek V3 | Janus Pro | Veo 2 | Gemini 2.5 Pro | Gemini 2.0 | Gemma 3 | Claude Sonnet 3.7 | Claude 3.5 Sonnet | Phi 4 | Phi 3.5 | Mistral Small 3.1 | Mistral NeMo | Mistral-7b | Bedrock | Vertex AI | Qwen QwQ 32B | Qwen 2 | Qwen 2.5 VL | Qwen Chat | Grok 3

Data Science Tools and Techniques

Python | R | SQL | Jupyter Notebooks | TensorFlow | Scikit-learn | PyTorch | Tableau | Apache Spark | Matplotlib | Seaborn | Pandas | Hadoop | Docker | Git | Keras | Apache Kafka | AWS | NLP | Random Forest | Computer Vision | Data Visualization | Data Exploration | Big Data | Common Machine Learning Algorithms | Machine Learning | Google Data Science Agent