

Machine Learning Engineer Interview Questions And Answers Guide.



Global Guideline.

<https://globalguideline.com/>



Machine Learning Engineer Job Interview Preparation Guide.

Question # 1

Tell us what is the difference between supervised and unsupervised machine learning?

Answer:-

Supervised learning requires training labeled data. For example, in order to do classification (a supervised learning task), you'll need to first label the data you'll use to train the model to classify data into your labeled groups. Unsupervised learning, in contrast, does not require labeling data explicitly.

[Read More Answers.](#)

Question # 2

Tell us when will you use classification over regression?

Answer:-

Classification is about identifying group membership while regression technique involves predicting a response. Both techniques are related to prediction, where classification predicts the belonging to a class whereas regression predicts the value from a continuous set. Classification technique is preferred over regression when the results of the model need to return the belongingness of data points in a dataset to specific explicit categories. (For instance, when you want to find out whether a name is male or female instead of just finding it how correlated they are with male and female names.

[Read More Answers.](#)

Question # 3

Can you name some feature extraction techniques used for dimensionality reduction?

Answer:-

- * Independent Component Analysis
- * Principal Component Analysis
- * Kernel Based Principal Component Analysis

[Read More Answers.](#)

Question # 4

Tell me how does deep learning contrast with other machine learning algorithms?

Answer:-

Deep learning is an approach to machine learning wherein the system learns the model as a neural network. If we're addressing the algorithms specifically, it should be noted that deep learning algorithms learn meaningful features on their own, without requiring any manual feature selection.

[Read More Answers.](#)

Question # 5

Tell us how do bias and variance play out in machine learning?

Answer:-

Both bias and variance are errors. Bias is an error due to flawed assumptions in the learning algorithm. Variance is an error resulting from too much complexity in the learning algorithm.

[Read More Answers.](#)

Question # 6

Tell me what are your favorite use cases of machine learning models?

Answer:-

The Quora thread above contains some examples, such as decision trees that categorize people into different tiers of intelligence based on IQ scores. Make sure that you have a few examples in mind and describe what resonated with you. It's important that you demonstrate an interest in how machine learning is implemented.

[Read More Answers.](#)

Question # 7



Machine Learning Engineer Interview Questions And Answers

Tell us which data visualization libraries do you use? What are your thoughts on the best data visualization tools?

Answer:-

What's important here is to define your views on how to properly visualize data and your personal preferences when it comes to tools. Popular tools include R's ggplot, Python's seaborn and matplotlib, and tools such as Plot.ly and Tableau.

[Read More Answers.](#)

Question # 8

How would you evaluate a logistic regression model?

Answer:-

A subsection of the question above. You have to demonstrate an understanding of what the typical goals of a logistic regression are (classification, prediction etc.) and bring up a few examples and use cases.

[Read More Answers.](#)

Question # 9

Tell us what's the difference between a generative and discriminative model?

Answer:-

A generative model will learn categories of data while a discriminative model will simply learn the distinction between different categories of data. Discriminative models will generally outperform generative models on classification tasks.

[Read More Answers.](#)

Question # 10

Tell me how a ROC curve works?

Answer:-

The ROC curve is a graphical representation of the contrast between true positive rates and the false positive rate at various thresholds. It's often used as a proxy for the trade-off between the sensitivity of the model (true positives) vs the fall-out or the probability it will trigger a false alarm (false positives).

[Read More Answers.](#)

Question # 11

Tell me what is the most frequent metric to assess model accuracy for classification problems?

Answer:-

Percent Correct Classification (PCC) measures the overall accuracy irrespective of the kind of errors that are made, all errors are considered to have same weight.

[Read More Answers.](#)

Question # 12

Explain me machine learning in to a layperson?

Answer:-

Machine learning is all about making decisions based on previous experience with a task with the intent of improving its performance. There are multiple examples that can be given to explain machine learning to a layperson -

* Imagine a curious kid who sticks his palm

* You have observed from your connections that obese people often tend to get heart diseases thus you make the decision that you will try to remain thin otherwise you might suffer from a heart disease. You have observed a ton of data and come up with a general rule of classification.

* You are playing blackjack and based on the sequence of cards you see, you decide whether to hit or to stay. In this case based on the previous information you have and by looking at what happens, you make a decision quickly.

[Read More Answers.](#)

Question # 13

Can you explain what is the difference between inductive machine learning and deductive machine learning?

Answer:-

In inductive machine learning, the model learns by examples from a set of observed instances to draw a generalized conclusion whereas in deductive learning the model first draws the conclusion and then the conclusion is drawn. Let's understand this with an example, for instance, if you have to explain to a kid that playing with fire can cause burns. There are two ways you can explain this to kids, you can show them training examples of various fire accidents or images with burnt people and label them as "Hazardous". In this case the kid will learn with the help of examples and not play with fire. This is referred to as Inductive machine learning. The other way is to let your kid play with fire and wait to see what happens. If the kid gets a burn they will learn not to play with fire and whenever they come across fire, they will avoid going near it. This is referred to as deductive learning.

[Read More Answers.](#)

Question # 14

Tell us how do classification and regression differ?

Answer:-

Classification predicts group or class membership. Regression involves predicting a response. Classification is the better technique when you need a more definite answer.

[Read More Answers.](#)

Question # 15



Explain me what is machine learning?

Answer:-

In answering this question, try to show your understand of the broad applications of machine learning, as well as how it fits into AI. Put it into your own words, but convey your understanding that machine learning is a form of AI that automates data analysis to enable computers to learn and adapt through experience to do specific tasks without explicit programming.

[Read More Answers.](#)

Question # 16

Tell me what are the last machine learning papers you've read?

Answer:-

Keeping up with the latest scientific literature on machine learning is a must if you want to demonstrate interest in a machine learning position. This overview of deep learning in Nature by the scions of deep learning themselves (from Hinton to Bengio to LeCun) can be a good reference paper and an overview of what's happening in deep learning - and the kind of paper you might want to cite.

[Read More Answers.](#)

Question # 17

Explain me a hash table?

Answer:-

A hash table is a data structure that produces an associative array. A key is mapped to certain values through the use of a hash function. They are often used for tasks such as database indexing.

[Read More Answers.](#)

Question # 18

Tell us what evaluation approaches would you work to gauge the effectiveness of a machine learning model?

Answer:-

You would first split the dataset into training and test sets, or perhaps use cross-validation techniques to further segment the dataset into composite sets of training and test sets within the data. You should then implement a choice selection of performance metrics. You could use measures such as the F1 score, the accuracy, and the confusion matrix. What's important here is to demonstrate that you understand the nuances of how a model is measured and how to choose the right performance measures for the right situations.

[Read More Answers.](#)

Question # 19

Explain me how would you handle an imbalanced dataset?

Answer:-

An imbalanced dataset is when you have, for example, a classification test and 90% of the data is in one class. That leads to problems: an accuracy of 90% can be skewed if you have no predictive power on the other category of data! Here are a few tactics to get over the hump:

- 1- Collect more data to even the imbalances in the dataset.
- 2- Resample the dataset to correct for imbalances.
- 3- Try a different algorithm altogether on your dataset.

What's important here is that you have a keen sense for what damage an unbalanced dataset can cause, and how to balance that.

[Read More Answers.](#)

Question # 20

Tell us how is a decision tree pruned?

Answer:-

Pruning is what happens in decision trees when branches that have weak predictive power are removed in order to reduce the complexity of the model and increase the predictive accuracy of a decision tree model. Pruning can happen bottom-up and top-down, with approaches such as reduced error pruning and cost complexity pruning.

Reduced error pruning is perhaps the simplest version: replace each node. If it doesn't decrease predictive accuracy, keep it pruned. While simple, this heuristic actually comes pretty close to an approach that would optimize for maximum accuracy.

[Read More Answers.](#)

Question # 21

Explain me what's the trade-off between bias and variance?

Answer:-

Bias is error due to erroneous or overly simplistic assumptions in the learning algorithm you're using. This can lead to the model underfitting your data, making it hard for it to have high predictive accuracy and for you to generalize your knowledge from the training set to the test set.

Variance is error due to too much complexity in the learning algorithm you're using. This leads to the algorithm being highly sensitive to high degrees of variation in your training data, which can lead your model to overfit the data. You'll be carrying too much noise from your training data for your model to be very useful for your test data.

The bias-variance decomposition essentially decomposes the learning error from any algorithm by adding the bias, the variance and a bit of irreducible error due to noise in the underlying dataset. Essentially, if you make the model more complex and add more variables, you'll lose bias but gain some variance - in order to get the optimally reduced amount of error, you'll have to tradeoff bias and variance. You don't want either high bias or high variance in your model.

[Read More Answers.](#)

Question # 22



Explain me what cross-validation technique would you use on a time series dataset?

Answer:-

Instead of using standard k-folds cross-validation, you have to pay attention to the fact that a time series is not randomly distributed data - it is inherently ordered by chronological order. If a pattern emerges in later time periods for example, your model may still pick up on it even if that effect doesn't hold in earlier years! You'll want to do something like forward chaining where you'll be able to model on past data then look at forward-facing data.

fold 1 : training [1], test [2]
fold 2 : training [1 2], test [3]
fold 3 : training [1 2 3], test [4]
fold 4 : training [1 2 3 4], test [5]
fold 5 : training [1 2 3 4 5], test [6]

[Read More Answers.](#)

Question # 23

Tell us which one would you prefer to choose - model accuracy or model performance?

Answer:-

Model accuracy is just a subset of model performance but is not the be-all and end-all of model performance. This question is asked to test your knowledge on how well you can make a perfect balance between model accuracy and model performance.

[Read More Answers.](#)

Question # 24

Tell us how will you know which machine learning algorithm to choose for your classification problem?

Answer:-

If accuracy is a major concern for you when deciding on a machine learning algorithm then the best way to go about it is test a couple of different ones (by trying different parameters within each algorithm) and choose the best one by cross-validation. A general rule of thumb to choose a good enough machine learning algorithm for your classification problem is based on how large your training set is. If the training set is small then using low variance/high bias classifiers like Naive Bayes is advantageous over high variance/low bias classifiers like k-nearest neighbour algorithms as it might overfit the model. High variance/low bias classifiers tend to win when the training set grows in size.

[Read More Answers.](#)

Question # 25

Please explain what is deep learning?

Answer:-

This might or might not apply to the job you're going after, but your answer will help to show you know more than just the technical aspects of machine learning. Deep learning is a subset of machine learning. It refers to using multi-layered neural networks to process data in increasingly complex ways, enabling the software to train itself to perform tasks like speech and image recognition through exposure to these vast amounts of data. Thus the machine undergoes continual improvement in the ability to recognize and process information. Layers of neural networks stacked on top of each for use in deep learning are called deep neural networks.

[Read More Answers.](#)

Question # 26

Do you know what's the "kernel trick" and how is it useful?

Answer:-

The Kernel trick involves kernel functions that can enable in higher-dimension spaces without explicitly calculating the coordinates of points within that dimension: instead, kernel functions compute the inner products between the images of all pairs of data in a feature space. This allows them the very useful attribute of calculating the coordinates of higher dimensions while being computationally cheaper than the explicit calculation of said coordinates. Many algorithms can be expressed in terms of inner products. Using the kernel trick enables us effectively run algorithms in a high-dimensional space with lower-dimensional data.

[Read More Answers.](#)

Question # 27

Please explain what is deep learning, and how does it contrast with other machine learning algorithms?

Answer:-

Deep learning is a subset of machine learning that is concerned with neural networks: how to use backpropagation and certain principles from neuroscience to more accurately model large sets of unlabelled or semi-structured data. In that sense, deep learning represents an unsupervised learning algorithm that learns representations of data through the use of neural nets.

[Read More Answers.](#)

Question # 28

Explain me what is Bayes' Theorem? How is it useful in a machine learning context?

Answer:-

Bayes' Theorem gives you the posterior probability of an event given what is known as prior knowledge. Mathematically, it's expressed as the true positive rate of a condition sample divided by the sum of the false positive rate of the population and the true positive rate of a condition. Say you had a 60% chance of actually having the flu after a flu test, but out of people who had the flu, the test will be false 50% of the time, and the overall population only has a 5% chance of having the flu. Would you actually have a 60% chance of having the flu after having a positive test?

Bayes' Theorem says no. It says that you have a $(.6 * 0.05) / (.6 * 0.05 + (.5 * 0.95))$ (True Positive Rate of a Condition Sample) / $(.6 * 0.05 + (.5 * 0.95))$ (True Positive Rate of a Condition Sample) + (False Positive Rate of a Population) = 0.0594 or 5.94% chance of getting a flu.

[Read More Answers.](#)

Question # 29



Do you know which is more important to you- model accuracy, or model performance?

Answer:-

This question tests your grasp of the nuances of machine learning model performance! Machine learning interview questions often look towards the details. There are models with higher accuracy that can perform worse in predictive power - how does that make sense? Well, it has everything to do with how model accuracy is only a subset of model performance, and at that, a sometimes misleading one. For example, if you wanted to detect fraud in a massive dataset with a sample of millions, a more accurate model would most likely predict no fraud at all if only a vast minority of cases were fraud. However, this would be useless for a predictive model - a model designed to find fraud that asserted there was no fraud at all! Questions like this help you demonstrate that you understand model accuracy isn't the be-all and end-all of model performance.

[Read More Answers.](#)

Question # 30

Tell us how do you ensure you're not overfitting with a model?

Answer:-

This is a simple restatement of a fundamental problem in machine learning: the possibility of overfitting training data and carrying the noise of that data through to the test set, thereby providing inaccurate generalizations.

There are three main methods to avoid overfitting:

- 1- Keep the model simpler: reduce variance by taking into account fewer variables and parameters, thereby removing some of the noise in the training data.
- 2- Use cross-validation techniques such as k-folds cross-validation.
- 3- Use regularization techniques such as LASSO that penalize certain model parameters if they're likely to cause overfitting.

[Read More Answers.](#)

Question # 31

Tell us what do you think of our current data process?

Answer:-

This kind of question requires you to listen carefully and impart feedback in a manner that is constructive and insightful. Your interviewer is trying to gauge if you'd be a valuable member of their team and whether you grasp the nuances of why certain things are set the way they are in the company's data process based on company- or industry-specific conditions. They're trying to see if you can be an intellectual peer. Act accordingly.

[Read More Answers.](#)

Question # 32

Tell us what is your training in machine learning and what types of hands-on experience do you have?

Answer:-

Your answer to this question will depend on your training in machine learning. Be sure to emphasize any direct projects you've completed as part of your education. Don't fail to mention any additional experience that you have including certifications and how they have prepared you for your role in the machine learning field.

[Read More Answers.](#)

Question # 33

Tell us why is Naive Bayes machine learning algorithm naive?

Answer:-

Naive Bayes machine learning algorithm is considered Naive because the assumptions the algorithm makes are virtually impossible to find in real-life data. Conditional probability is calculated as a pure product of individual probabilities of components. This means that the algorithm assumes the presence or absence of a specific feature of a class is not related to the presence or absence of any other feature (absolute independence of features), given the class variable. For instance, a fruit may be considered to be a banana if it is yellow, long and about 5 inches in length. However, if these features depend on each other or are based on the existence of other features, a naive Bayes classifier will assume all these properties to contribute independently to the probability that this fruit is a banana. Assuming that all features in a given dataset are equally important and independent rarely exists in the real-world scenario.

[Read More Answers.](#)

Question # 34

Tell me how is KNN different from k-means clustering?

Answer:-

K-Nearest Neighbors is a supervised classification algorithm, while k-means clustering is an unsupervised clustering algorithm. While the mechanisms may seem similar at first, what this really means is that in order for K-Nearest Neighbors to work, you need labeled data you want to classify an unlabeled point into (thus the nearest neighbor part). K-means clustering requires only a set of unlabeled points and a threshold: the algorithm will take unlabeled points and gradually learn how to cluster them into groups by computing the mean of the distance between different points.

The critical difference here is that KNN needs labeled points and is thus supervised learning, while k-means doesn't - and is thus unsupervised learning.

[Read More Answers.](#)

Question # 35

Explain me what's your favorite algorithm, and can you explain it to me in less than a minute?

Answer:-

This type of question tests your understanding of how to communicate complex and technical nuances with poise and the ability to summarize quickly and efficiently. Make sure you have a choice and make sure you can explain different algorithms so simply and effectively that a five-year-old could grasp the basics!

[Read More Answers.](#)

Question # 36

Can you list some use cases where classification machine learning algorithms can be used?

**Answer:-**

- * Natural language processing (Best example for this is Spoken Language Understanding)
- * Market Segmentation
- * Text Categorization (Spam Filtering)
- * Bioinformatics (Classifying proteins according to their function)
- * Fraud Detection
- * Face detection

[Read More Answers.](#)

Question # 37

Tell me what is precision and recall?

Answer:-

Recall is also known as the true positive rate: the amount of positives your model claims compared to the actual number of positives there are throughout the data. Precision is also known as the positive predictive value, and it is a measure of the amount of accurate positives your model claims compared to the number of positives it actually claims. It can be easier to think of recall and precision in the context of a case where you've predicted that there were 10 apples and 5 oranges in a case of 10 apples. You'd have perfect recall (there are actually 10 apples, and you predicted there would be 10) but 66.7% precision because out of the 15 events you predicted, only 10 (the apples) are correct.

[Read More Answers.](#)

Question # 38

Tell us an example where ensemble techniques might be useful?

Answer:-

Ensemble techniques use a combination of learning algorithms to optimize better predictive performance. They typically reduce overfitting in models and make the model more robust (unlikely to be influenced by small changes in the training data).

You could list some examples of ensemble methods, from bagging to boosting to a "bucket of models" method and demonstrate how they could increase predictive power.

[Read More Answers.](#)

Question # 39

Tell us how can we use your machine learning skills to generate revenue?

Answer:-

This is a tricky question. The ideal answer would demonstrate knowledge of what drives the business and how your skills could relate. For example, if you were interviewing for music-streaming startup Spotify, you could remark that your skills at developing a better recommendation model would increase user retention, which would then increase revenue in the long run.

The startup metrics Slideshare linked above will help you understand exactly what performance indicators are important for startups and tech companies as they think about revenue and growth.

[Read More Answers.](#)

Question # 40

Tell me what is supervised versus unsupervised learning?

Answer:-

Supervised learning is a process of machine learning in which outputs are fed back into a computer for the software to learn from for more accurate results the next time. With supervised learning, the "machine" receives initial training to start. In contrast, unsupervised learning means a computer will learn without initial training.

[Read More Answers.](#)

Question # 41

Tell me how much data will you allocate for your training, validation and test sets?

Answer:-

There is no to the point answer to this question but there needs to be a balance/equilibrium when allocating data for training, validation and test sets.

If you make the training set too small, then the actual model parameters might have high variance. Also, if the test set is too small, there are chances of unreliable estimation of model performance. A general thumb rule to follow is to use 80: 20 train/test split. After this the training set can be further split into validation sets.

[Read More Answers.](#)

Question # 42

Tell us why is "Naive" Bayes naive?

Answer:-

Despite its practical applications, especially in text mining, Naive Bayes is considered "Naive" because it makes an assumption that is virtually impossible to see in real-life data: the conditional probability is calculated as the pure product of the individual probabilities of components. This implies the absolute independence of features - a condition probably never met in real life.

As a Quora commenter put it whimsically, a Naive Bayes classifier that figured out that you liked pickles and ice cream would probably naively recommend you a pickle ice cream.

[Read More Answers.](#)

Question # 43

Tell me do you have experience with Spark or big data tools for machine learning?

Answer:-



Machine Learning Engineer Interview Questions And Answers

You'll want to get familiar with the meaning of big data for different companies and the different tools they'll want. Spark is the big data tool most in demand now, able to handle immense datasets with speed. Be honest if you don't have experience with the tools demanded, but also take a look at job descriptions and see what tools pop up: you'll want to invest in familiarizing yourself with them.

[Read More Answers.](#)

Question # 44

Tell us what is decision tree classification?

Answer:-

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

[Read More Answers.](#)

Question # 45

Tell us what kind of problems does regularization solve?

Answer:-

Regularization is used to address overfitting problems as it penalizes the loss function by adding a multiple of an L1 (LASSO) or an L2 (Ridge) norm of your weights vector w .

[Read More Answers.](#)

Question # 46

Tell us how do deductive and inductive machine learning differ?

Answer:-

Deductive machine learning starts with a conclusion, then learns by deducing what is right or wrong about that conclusion. Inductive machine learning starts with examples from which to draw conclusions.

[Read More Answers.](#)

Question # 47

Tell us when should you use classification over regression?

Answer:-

Classification produces discrete values and dataset to strict categories, while regression gives you continuous results that allow you to better distinguish differences between individual points. You would use classification over regression if you wanted your results to reflect the belongingness of data points in your dataset to certain explicit categories (ex: If you wanted to know whether a name was male or female rather than just how correlated they were with male and female names.)

[Read More Answers.](#)

Question # 48

Tell us what's the difference between Type I and Type II error?

Answer:-

Don't think that this is a trick question! Many machine learning interview questions will be an attempt to lob basic questions at you just to make sure you're on top of your game and you've prepared all of your bases.

Type I error is a false positive, while Type II error is a false negative. Briefly stated, Type I error means claiming something has happened when it hasn't, while Type II error means that you claim nothing is happening when in fact something is.

A clever way to think about this is to think of Type I error as telling a man he is pregnant, while Type II error means you tell a pregnant woman she isn't carrying a baby.

[Read More Answers.](#)

Question # 49

Explain how do you think Google is training data for self-driving cars?

Answer:-

Machine learning interview questions like this one really test your knowledge of different machine learning methods, and your inventiveness if you don't know the answer. Google is currently using recaptcha to source labelled data on storefronts and traffic signs. They are also building on training data collected by Sebastian Thrun at GoogleX - some of which was obtained by his grad students driving buggies on desert dunes!

[Read More Answers.](#)

Question # 50

Tell us what are some differences between a linked list and an array?

Answer:-

An array is an ordered collection of objects. A linked list is a series of objects with pointers that direct how to process them sequentially. An array assumes that every element has the same size, unlike the linked list. A linked list can more easily grow organically: an array has to be pre-defined or re-defined for organic growth. Shuffling a linked list involves changing which points direct where - meanwhile, shuffling an array is more complex and takes more memory.

[Read More Answers.](#)

Question # 51

Tell us what's a Fourier transform?

Answer:-



A Fourier transform is a generic method to decompose generic functions into a superposition of symmetric functions. Or as this more intuitive tutorial puts it, given a smoothie, it's how we find the recipe. The Fourier transform finds the set of cycle speeds, amplitudes and phases to match any time signal. A Fourier transform converts a signal from time to frequency domain - it's a very common way to extract features from audio signals or other time series such as sensor data.

[Read More Answers.](#)

Question # 52

Tell us how would you approach the "Netflix Prize" competition?

Answer:-

The Netflix Prize was a famed competition where Netflix offered \$1,000,000 for a better collaborative filtering algorithm. The team that won called BellKor had a 10% improvement and used an ensemble of different methods to win. Some familiarity with the case and its solution will help demonstrate you've paid attention to machine learning for a while.

[Read More Answers.](#)

Question # 53

Tell us which do you think is more important: model accuracy or model performance?

Answer:-

While both accuracy and performance are of course important, and subjective to the specific application you're building, accuracy is more important in general. If your machine learning application provides inaccurate information, it doesn't matter how quickly it does it.

[Read More Answers.](#)

Question # 54

Tell us do you have research experience in machine learning?

Answer:-

Related to the last point, most organizations hiring for machine learning positions will look for your formal experience in the field. Research papers, co-authored or supervised by leaders in the field, can make the difference between you being hired and not. Make sure you have a summary of your research experience and papers ready - and an explanation for your background and lack of formal research experience if you don't.

[Read More Answers.](#)

Question # 55

Tell us how do you handle missing or corrupted data in a dataset?

Answer:-

You could find missing/corrupted data in a dataset and either drop those rows or columns, or decide to replace them with another value. In Pandas, there are two very useful methods: `isnull()` and `dropna()` that will help you find columns of data with missing or corrupted data and drop those values. If you want to fill the invalid values with a placeholder value (for example, 0), you could use the `fillna()` method.

[Read More Answers.](#)

Question # 56

Tell me what is the difference between bias and variance?

Answer:-

Bias comes as a consequence of a model underfitting some set of data, whereas variance arises as the result of overfitting some set of data.

[Read More Answers.](#)

Question # 57

Can you pick an algorithm. Write the psuedo-code for a parallel implementation?

Answer:-

This kind of question demonstrates your ability to think in parallelism and how you could handle concurrency in programming implementations dealing with big data. Take a look at pseudocode frameworks such as Peril-L and visualization tools such as Web Sequence Diagrams to help you demonstrate your ability to write code that reflects parallelism.

[Read More Answers.](#)

Question # 58

What is the difference between L1 and L2 regularization?

Answer:-

L2 regularization tends to spread error among all the terms, while L1 is more binary/sparse, with many variables either being assigned a 1 or 0 in weighting. L1 corresponds to setting a Laplacean prior on the terms, while L2 corresponds to a Gaussian prior.

[Read More Answers.](#)

Question # 59

Tell us where do you usually source datasets?

Answer:-

Machine learning interview questions like these try to get at the heart of your machine learning interest. Somebody who is truly passionate about machine learning will have gone off and done side projects on their own, and have a good idea of what great datasets are out there. If you're missing any, check out Quandl for economic and financial data, and Kaggle's Datasets collection for another great list.



[Read More Answers.](#)

Question # 60

Tell me what is a recommendation system?

Answer:-

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system: It's an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

[Read More Answers.](#)

Question # 61

Explain what are some methods of reducing dimensionality?

Answer:-

You can reduce dimensionality by combining features with feature engineering, removing collinear features, or using algorithmic dimensionality reduction.

[Read More Answers.](#)

Question # 62

Tell us what's the F1 score? How would you use it?

Answer:-

The F1 score is a measure of a model's performance. It is a weighted average of the precision and recall of a model, with results tending to 1 being the best, and those tending to 0 being the worst. You would use it in classification tests where true negatives don't matter much.

[Read More Answers.](#)

Question # 63

Tell me how do you choose an algorithm for a classification problem?

Answer:-

The answer depends on the degree of accuracy needed and the size of the training set. If you have a small training set, you can use a low variance/high bias classifier. If your training set is large, you will want to choose a high variance/low bias classifier.

[Read More Answers.](#)

Question # 64

Tell me how would you implement a recommendation system for our company's users?

Answer:-

A lot of machine learning interview questions of this type will involve implementation of machine learning models to a company's problems. You'll have to research the company and its industry in-depth, especially the revenue drivers the company has, and the types of users the company takes on in the context of the industry it's in.

[Read More Answers.](#)

Question # 65

Do you know what is kernel SVM?

Answer:-

Kernel SVM is the abbreviated version of kernel support vector machine. Kernel methods are a class of algorithms for pattern analysis and the most common one is the kernel SVM.

[Read More Answers.](#)

Engineering Most Popular Interview Topics.

- 1 : [Civil Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 2 : [Mechanical Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 3 : [Electrical Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 4 : [Chemical Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 5 : [Automobile Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 6 : [Electronics Communications Frequently Asked Interview Questions and Answers Guide.](#)
- 7 : [Instrumentation Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 8 : [Marine Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 9 : [Industrial Engineering Frequently Asked Interview Questions and Answers Guide.](#)
- 10 : [RF Engineer Frequently Asked Interview Questions and Answers Guide.](#)

About Global Guideline.

Global Guideline is a platform to develop your own skills with thousands of job interview questions and web tutorials for fresher's and experienced candidates. These interview questions and web tutorials will help you strengthen your technical skills, prepare for the interviews and quickly revise the concepts. Global Guideline invite you to unlock your potentials with thousands of [Interview Questions with Answers](#) and much more. Learn the most common technologies at Global Guideline. We will help you to explore the resources of the World Wide Web and develop your own skills from the basics to the advanced. Here you will learn anything quite easily and you will really enjoy while learning. Global Guideline will help you to become a professional and Expert, well prepared for the future.

* This PDF was generated from <https://GlobalGuideline.com> at **November 29th, 2023**

* If any answer or question is incorrect or inappropriate or you have correct answer or you found any problem in this document then don't hesitate feel free and [e-mail us](#) we will fix it.

You can follow us on FaceBook for latest Jobs, Updates and other interviews material.
www.facebook.com/InterviewQuestionsAnswers

Follow us on Twitter for latest Jobs and interview preparation guides
<https://twitter.com/InterviewGuide>

Best Of Luck.

Global Guideline Team
<https://GlobalGuideline.com>
Info@globalguideline.com