

[Home](#) > [Blog](#) > [Machine Learning](#)

Top 30 Machine Learning Interview Questions For 2025

Prepare for your interview with this comprehensive guide to machine learning questions, covering everything from basic concepts and algorithms to advanced and role-specific topics.

Updated Jan 11, 2025 · 15 min read

**Abid Ali Awan**

Certified data scientist, passionate about building ML apps, blogging on data science, and editing.

TOPICS

[Machine Learning](#)[Data Science](#)

In this article, we've outlined some of the most common machine learning interview questions you might encounter when applying for a role in the industry or interviewing a candidate. Practicing these questions and preparing answers can help you make sure the interview goes smoothly.

Develop AI Applications

Learn to build AI applications using the OpenAI API.

[Start Upskilling For Free](#)[Download Now](#)

ask these questions to assess the technical knowledge of the candidate.

What is semi-supervised machine learning?

Semi-supervised learning is the blend of supervised and unsupervised learning. The algorithm is trained on a mix of labeled and unlabeled data. Generally, it is utilized when we have a very small labeled dataset and a large unlabeled dataset.

In simple terms, the unsupervised algorithm is used to create clusters and by using existing labeled data to label the rest of the unlabelled data. A Semi-supervised algorithm assumes continuity assumption, cluster assumption, and manifold assumption.

It is generally used to save the cost of acquiring labeled data. For example, protein sequence classification, automatic speech recognition, and self-driving cars.



How do you choose which algorithm to use for a dataset?

Apart from the dataset, you need a business use case or application requirements. You can apply supervised and unsupervised learning to the same data.

Generally:

- Supervised learning algorithms require labeled data.
 - Regression algorithms require continuous numerical targets
 - Classification algorithms require categorical targets
- Unsupervised learning algorithms require unlabeled data.
- Semi-supervised learning requires the combination of labeled and unlabeled datasets.
- Reinforcement learning algorithms require environment, agent, state, and reward data.

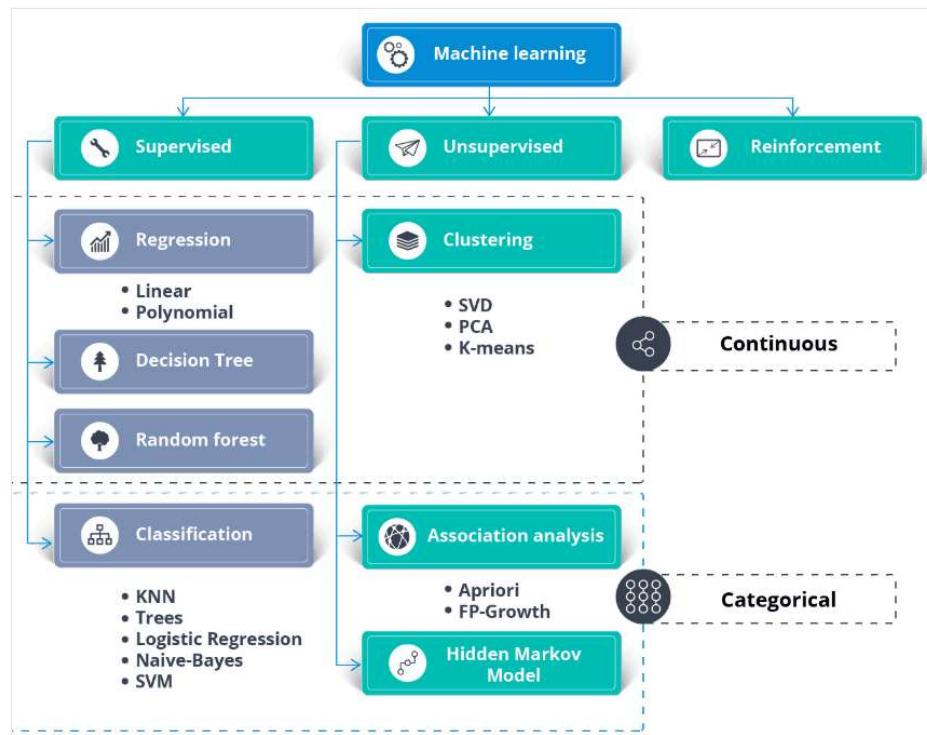


Image from [thecleverprogrammer](#)

Learn the [machine learning fundamentals](#) by taking our course.

Explain the K Nearest Neighbor algorithm.

The K Nearest Neighbor (KNN) is a supervised learning classifier. It uses proximity to classify labels or predict the grouping of individual data points. We can use it for regression and classification. KNN algorithm is non-parametric, meaning it doesn't make an underlying assumption of data distribution.

In the KNN classifier:

- We find K-neighbors nearest to the white point. In the example below, we chose k=5.
- To find the five nearest neighbors, we calculate the euclidean distance between the white point and the others. Then, we chose the 5 points closest to the white point.
- There are three red and two green points at K=5. Since the red has a majority, we assign a red label to it.

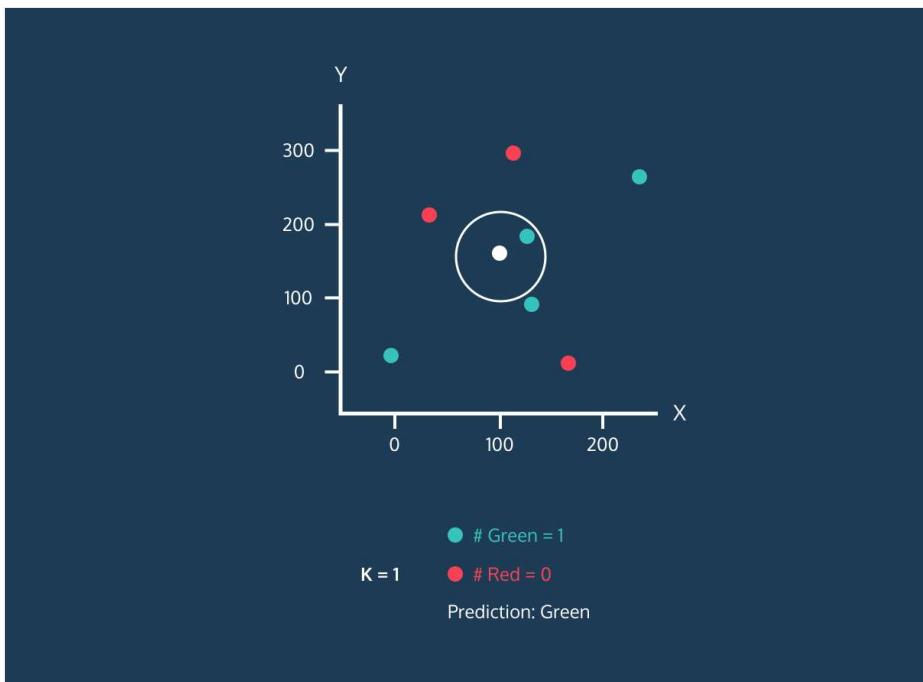


Image from [Codesigner's Dev Story](#)

Learn all about [supervised learning classification and regression models](#) by taking a short course.

What is feature importance in machine learning, and how do you determine it?

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. It plays a critical role in understanding the data's underlying structure, the behavior of the model, and making the model more interpretable.

There are several methods to determine feature importance:

1. **Model-based Importance:** Certain algorithms like Decision Trees and Random Forests provide built-in methods to evaluate feature importance. For example, Random Forests calculate the decrease in node impurity weighted by the probability of reaching that node, averaged over all trees.
2. **Permutation Importance:** This involves shuffling individual variables in the validation set and observing the effect on model performance. A significant decrease in model performance indicates high importance.
3. **SHAP (SHapley Additive exPlanations):** This approach uses game theory to measure the contribution of each feature to the prediction in a complex model. [SHAP values](#) provide a deep insight into the model's behavior and are particularly useful for complex models like gradient boosting machines or neural networks.
4. **Correlation Coefficients:** Simple statistical measures like Pearson or Spearman correlation can provide insights into the linear relationship between each feature and the target variable.

Understanding feature importance is crucial for model optimization, reducing overfitting by removing non-informative features, and improving model interpretability, especially in domains where understanding the model's decision process is critical.

What is overfitting in machine learning, and how can it be avoided?

Overfitting occurs when a model performs well on training data but fails to generalize to unseen data because it has memorized the training data instead of learning the underlying patterns. It can be avoided by:

- Using cross-validation techniques.
- Applying regularization (L1, L2).
- Reducing model complexity.
- Collecting more training data or using data augmentation.

What is the difference between supervised and unsupervised learning?

- **Supervised Learning:** The model is trained on labeled data, where the target variable is known. Examples include classification and regression.
- **Unsupervised Learning:** The model is trained on unlabeled data to find patterns or groupings. Examples include clustering and dimensionality reduction.

What is a confusion matrix, and why is it useful?

A confusion matrix is a table used to evaluate the performance of a classification model. It shows the counts of true positives, true negatives, false positives, and false negatives. It is useful for calculating metrics like accuracy, precision, recall, and F1 score.

What is the difference between parametric and non-parametric models?

Parametric Models: These make assumptions about the data's underlying distribution and have a fixed number of parameters (e.g., Linear Regression).

Non-Parametric Models: These make no assumptions about data distribution and can adapt to complexity as more data is added (e.g., K-Nearest Neighbors).

What is the bias-variance tradeoff in machine learning?

The bias-variance tradeoff refers to the balance between a model's ability to capture complex patterns (low bias) and its sensitivity to fluctuations in the training data (low variance). A good model achieves a balance by minimizing both bias and variance to avoid underfitting and overfitting.

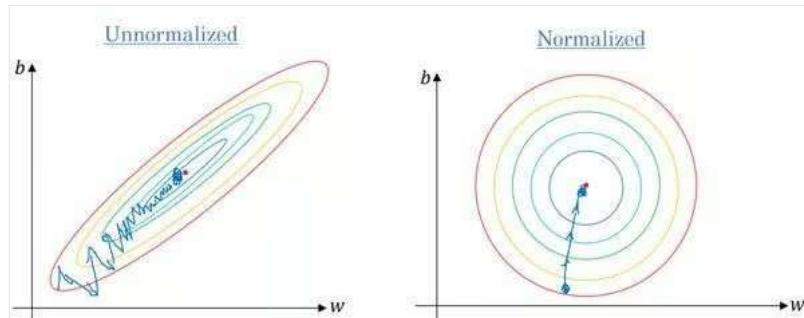
Machine Learning Technical Interview Questions

The technical interview session is more about assessing your knowledge about processes and how well you are equipped to handle uncertainty. The hiring manager will ask machine learning interview questions about data processing, model training and validation, and advanced algorithms.

Is it true that we need to scale our feature values when they vary greatly?

Yes. Most of the algorithms use Euclidean distance between data points, and if the feature value varies greatly, the results will be quite different. In most cases, outliers cause machine learning models to perform worse on the test dataset.

We also use feature scaling to reduce convergence time. It will take longer for gradient descent to reach local minima when features are not normalized.

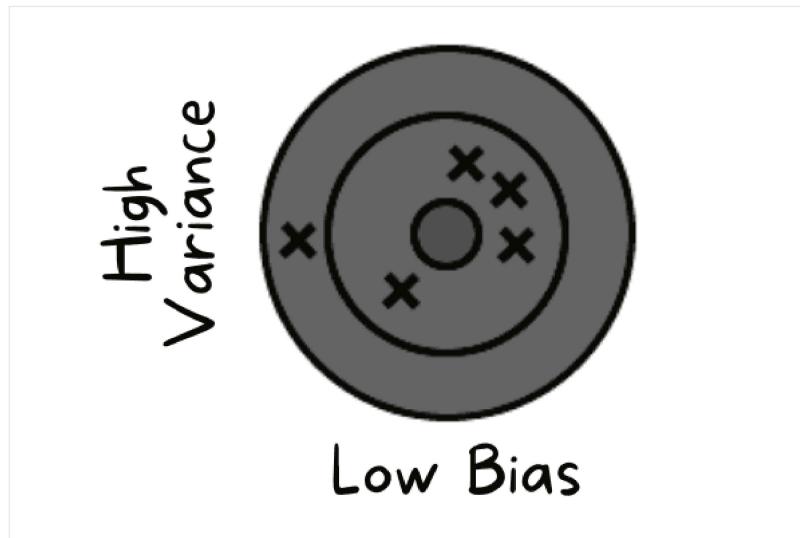


Gradient without and with scaling | [Quora](#)

Feature engineering skills are in high demand. You can learn everything about the subject by taking a DataCamp course, such as [Feature Engineering for Machine Learning in Python](#).

The model you have trained has a low bias and high variance. How would you deal with it?

Low bias occurs when the model is predicting values close to the actual value. It is mimicking the training dataset. The model has no generalization which means if the model is tested on unseen data, it will give poor results.



Low bias and high variance | Author

To fix these issues, we will use bagging algorithms as it divides a data set into subsets using randomized sampling. Then, we generate sets of models using these samples with a single algorithm. After that, we combine the model prediction using voting classification or averaging.

For high variance, we can use regularization techniques. It penalized higher model coefficients to lower model complexity. Furthermore, we can select the top features from the feature importance graph and train the model.

Which cross-validation technique would you suggest for a time-series dataset and why?

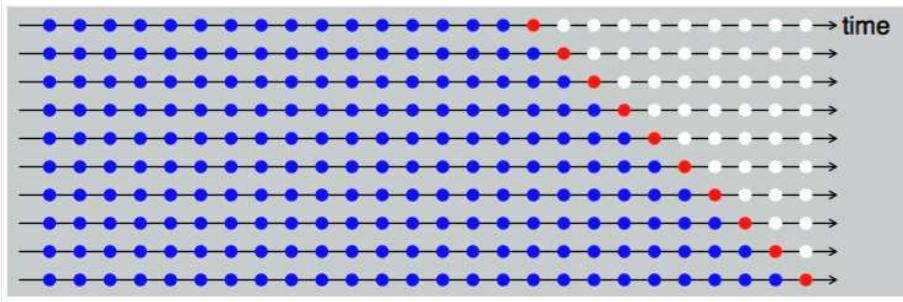
Cross-validation is used to evaluate model performance robustly and prevent overfitting. Generally, cross-validation techniques randomly pick samples from the data and split them into train and test data sets. The number of splits is based on the K value.

For example, if the K = 5, there will be four folds for the train and one for the test. It will repeat five times to measure the model performed on separate folds.

We cannot do it with a time series dataset because it doesn't make sense to use the value from the future to forecast the value of the past. There is a temporal dependency between observations, and we can only split the data in one direction so that the values of the test dataset are after the training set.

The diagram shows that time series data k fold split is unidirectional. The blue points are the training set, the red point is the test set, and the white is unused data. As we can observe with every iteration, we are moving forward with the training set while the test set remains in front of the training set, not randomly selected.

Time series cross-validation



Time series cross validation | [UC Business Analytics R Programming Guide](#)

Learn about time series data manipulation, analysis, visualization, and modeling by taking [Time Series with Python](#).

Computer Vision Engineering Interview Questions

Most machine learning jobs offered on LinkedIn, Glassdoor, and Indeed are role specific. As such, during the interview, they will focus on role-specific questions. For the computer vision engineering role the hiring manager will focus on image processing questions.

Why can the inputs in computer vision problems get huge? Explain it with an example.

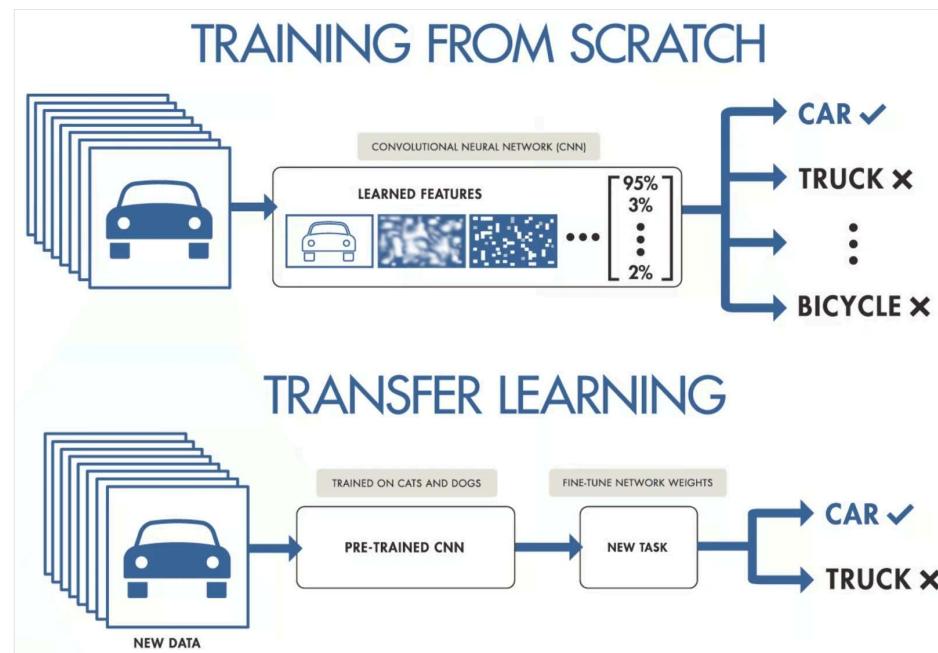
Imagine an image of 250 X 250 and a fully connected hidden first layer with 1000 hidden units. For this image, the input features are $250 \times 250 \times 3 = 187,500$, and the weight matrix at the first hidden layer will be $187,500 \times 1000$ dimensional matrix. These numbers are huge for storage and computation, and to combat this problem, we use convolution operations.

Learn image processing by taking a short [Image Processing in Python](#) course

When you have a small dataset, suggest a way to train a convolutional neural network.

If you do not have enough data to train a convolutional neural network, you can use transfer learning to train your model and get state-of-the-art results. You need a pre-trained model which was trained on a general but larger dataset. After that, you will fine-tune it on newer data by training the last layers of the models.

Transfer learning allows data scientists to train models on smaller data by using fewer resources, computing, and storage. You can find open-source pre-trained models for various use cases easily, and most of them have a commercial license which means you can use them to create your application.

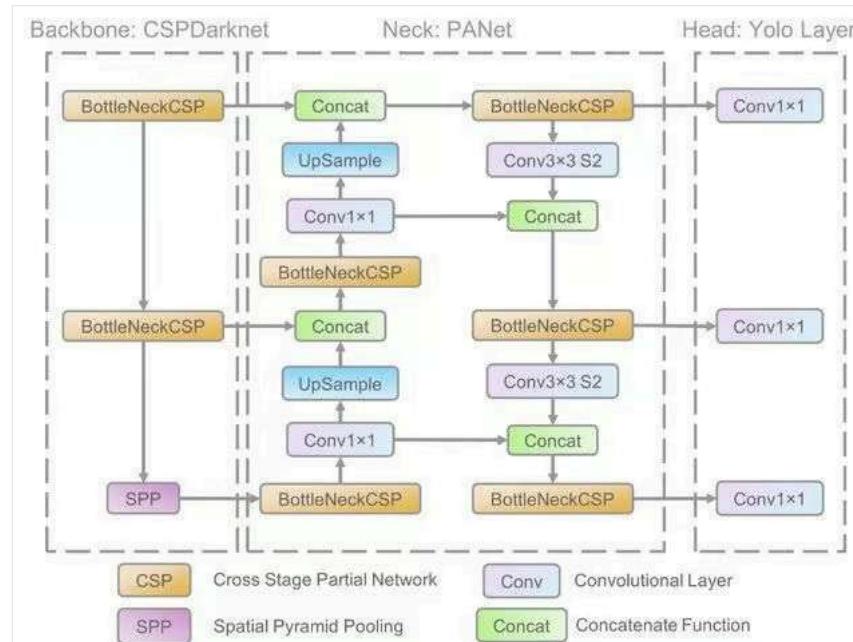


Transfer Learning by [purnasai gudikandula](#)

What is the state-of-the-art object detection algorithm YOLO?

YOLO is an object detection algorithm based on convolutional neural networks, and it can provide real-time results. The [YOLO algorithm](#) requires a single forward pass through CNN to recognize the object. It predicts both various class probabilities and boundary boxes.

The model was trained to detect various objects, and companies are using transfer learning to fine-tune it on new data for modern applications such as autonomous driving, wildlife preservation, and security.



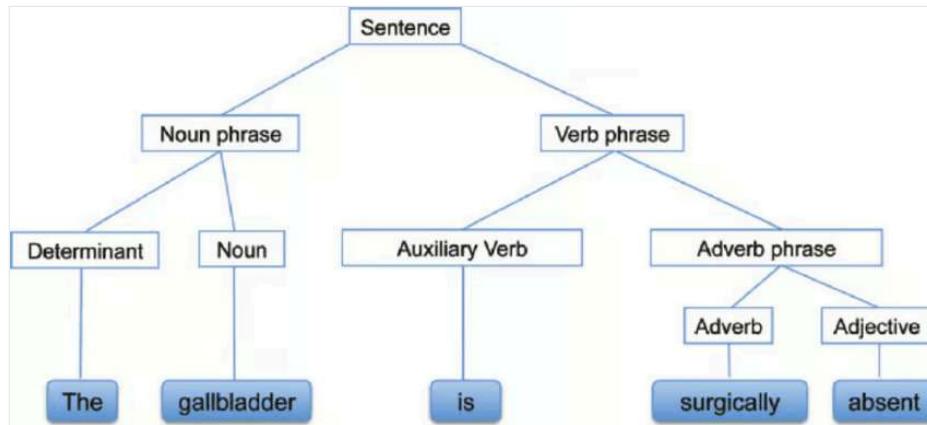
YOLO V5 model architecture | [researchgate](#)

NLP Engineering Interview Questions

What is Syntactic Analysis?

Syntactic Analysis, also known as Syntax analysis or Parsing, is a text analysis that tells us the logical meaning behind the sentence or part of the sentence. It focuses on the

relationship between words and the grammatical structure of sentences. You can also say that it is the processing of analyzing the natural language by using grammatical rules.



Syntactic Analysis | [researchgate](#)

What are Stemming and Lemmatization?

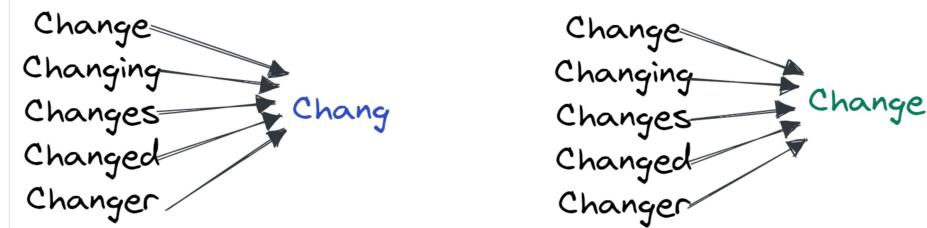
Stemming and lemmatization is a normalizing technique used to minimize the structural variation of words in a sentence.

Stemming removes the affixes added to the word and leaves it in base form. For example, Changing to Chang.

It is widely used by search engines for storage optimization. Instead of storing all the forms of the words, it only stores the stems.

Lemmatization converts the word into its lemma form. The output is the root word instead of the stem word. After lemmatization, we get the valid word that means something. For example, Changing to Change.

Stemming v/s Lemmatization



Stemming vs. Lemmatization | Author

How would you reduce the inference time of a trained transformer model?

It is the responsibility of machine learning engineers to optimize the model inference. Due to large language models, it has become more difficult to deploy models in production and reduce inference time to microseconds.

To improve inference time, we can use:

- GPU, TPU, or FPGA for acceleration.

- GPU with fp16 support
- Pruning to reduce parameters
- Knowledge distillation
- Hierarchical softmax or adaptive softmax
- Cache predictions
- Parallel/batch computing
- Reduce the model size

Learn the basics of NLP by completing [Natural Language Processing in Python](#) skill track.

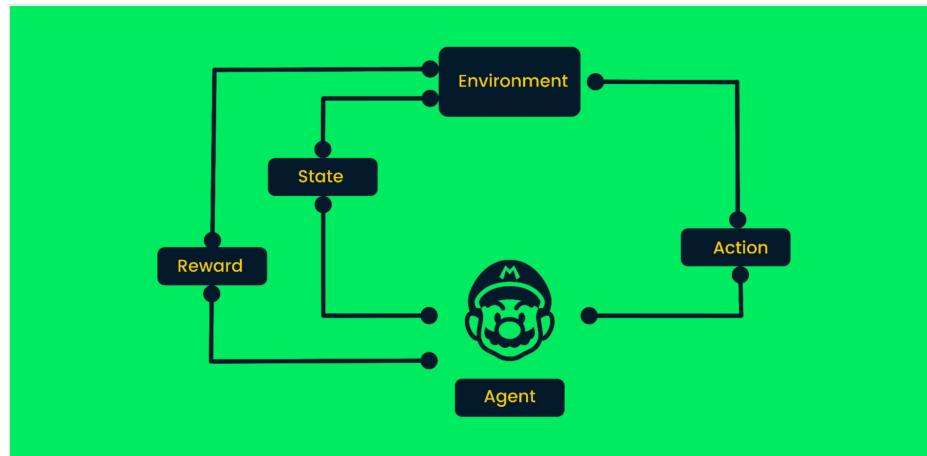
Reinforcement Learning Engineering Interview Questions

What are the steps involved in a typical Reinforcement Learning algorithm?

Reinforcement learning uses trial and error to reach goals. It is a goal-oriented algorithm and it learns from the environment by taking correct steps to maximize the cumulative reward.

In typical reinforcement learning:

1. At the start, the agent receives state zero from the environment
2. Based on the state, the agent will take an action
3. The state has changed, and the agent is at a new place in the environment.
4. The agent receives the reward if it has made the correct move.
5. The process will repeat until the agent has learned the best possible path to reach the goal by maximizing the cumulative rewards.



Reinforcement Learning Framework | Author

What is the difference between Off-Policy and On-Policy Learning?

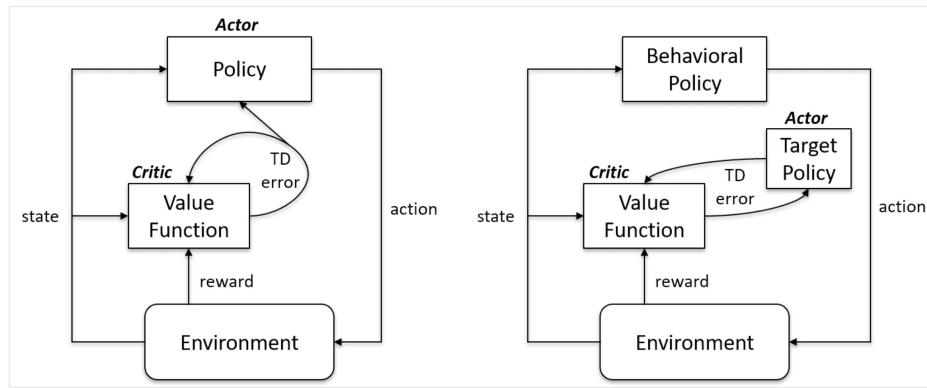
On-Policy learning algorithms evaluate and improve the same policy to act and update it. In other words, the policy that is used for updating and the policy that is used to take action are the same.

Target Policy == Behavior Policy

On-policy algorithms are Sarsa, Monte Carlo for On-Policy, Value Iteration, and Policy Iteration

Off-Policy Learning algorithms are completely different as the updated policy is different from the behavior policy. For example, in Q-learning, the agent learns from an optimal policy with the help of a greedy policy and takes action using other policies.

Target Policy \neq Behavior Policy



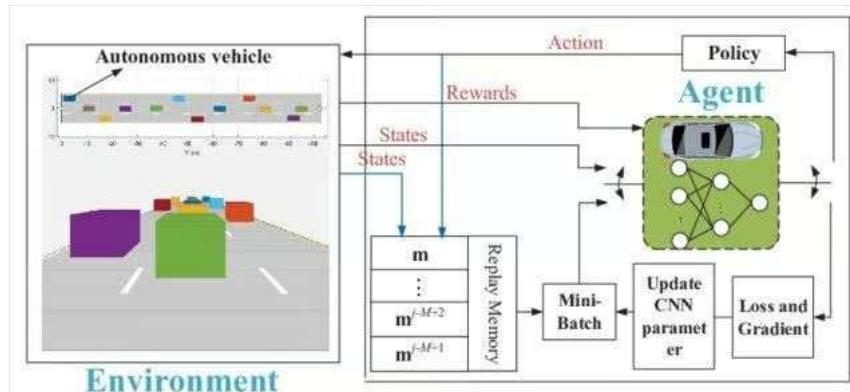
On-policy vs. Off-policy case | [Artificial Intelligence Stack Exchange](#)

Why do we need “Deep” Q learning?

Simple Q learning is great. It solves the problem on a smaller scale, but on a larger scale, it fails.

Imagine if the environment has 1000 states and 1000 actions per state. We will require a Q table of millions of cells. The game of chess and Go will require an even bigger table. This is where Deep Q-learning comes for the rescue.

It utilizes a neural network to approximate the Q value function. The neural networks recipe states as an input and outputs the Q-value of all possible actions.



Deep Q-network for autonomous driving | [researchgate](#)

FAANG Machine Learning Engineer Questions

Below, we've outlined some potential questions the interviewer might ask you at some of the top tech companies:

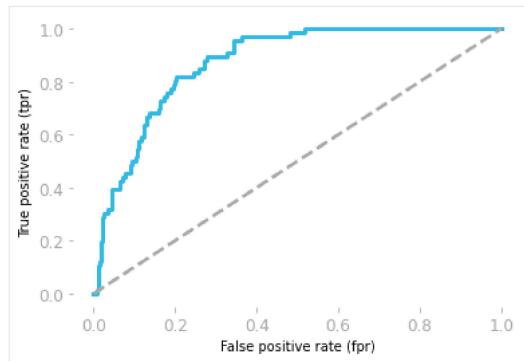
What is the interpretation of a ROC area under the curve?

Receiver operating characteristics (ROC) shows the trade-off between sensitivity and specificity.

- Sensitivity: it is the probability that the model predicts a positive outcome when the actual value is also positive.
- Specificity: it is the probability that the model predicts a negative outcome when the actual value is also negative.

The curve is plotted using the False positive rate ($FP/(TN + FP)$) and true positive rate ($TP/(TP + FN)$)

The area under the curve (AUC) shows the model performance. If the area under the ROC curve is 0.5, then our model is completely random. The model with AUC close to 1 is the better model.



ROC curve by Hadrien Jean

What are the methods of reducing dimensionality?

For dimensionality reduction, we can use feature selection or feature extraction methods.

Feature selection is a process of selecting optimal features and dropping irrelevant features. We use Filter, Wrapper, and Embedded methods to analyze feature importance and remove less important features to improve model performance.

Feature extraction transforms the space with multiple dimensions into fewer dimensions. No information is lost during the process, and it uses fewer resources to process the data. The most common extraction techniques are Linear discriminant analysis (LDA), Kernel PCA, and Quadratic discriminant analysis.

How do you find thresholds for a classifier?

In the case of a spam classifier, a logistics regression model will return the probability. We either use the probability of 0.8999 or convert it into class (Spam/Not Spam) using a threshold.

Usually, the threshold of a classifier is 0.5, but in some cases, we need to fine-tune it to improve the accuracy. The 0.5 threshold means that if the probability is equal to or above 0.5, it is spam, and if it is lower, then it is not spam.

To find the threshold, we can use Precision-Recall curves and ROC curves, grid search, and by manually changing the value to get a better CV.

Become a professional machine learning engineer by completing [Machine Learning Scientist with Python](#) career track.

What are the assumptions of linear regression?

Linear regression is used to understand the relation between features (X) and target (y). Before we train the model, we need to meet a few assumptions:

1. The residuals are independent
2. There is a linear relation between X independent variable and y dependent variable.
3. Constant residual variance at every level of X
4. The residuals are normally distributed.

Note: the residuals in linear regression are the difference between actual and predicted values.

Write a function `find_bigrams` to take a string and return a list of all bigrams.

During coding interviews, you will be asked about machine learning problems, but in some cases, they will assess your Python skills by asking you general coding questions. Become an expert Python programmer by taking the [Python Programmer](#) career track.

Creating a bigram function is quite easy. You need to use two loops with the `zip` function.

1. In bigram function, we are taking a list of the sentence as an input
2. Creating a loop to access a single sentence
3. Lowering and splitting the sentence into a list of words
4. Using `zip` to create a combination of the previous word and the next word
5. Appending the output to the result
6. Printing the results.

It is quite easy if you break down the problem and use `zip` functions.

```
def bigram(text_list:list):
    result = []
    for ls in text_list:
        words = ls.lower().split()
        for bi in zip(words, words[1:]):
            result.append(bi)
    return result
text = ["Data drives everything", "Get the skills you need for the future of work"]
print(bigram(text))
```

POWERED BY  datacamp

Results:

```
[('Data', 'drives'), ('drives', 'everything'), ('Get', 'the'), ('the', 'skills')]
```

POWERED BY  datacamp

What is the activation function in Machine Learning?

The activation function is a non-linear transformation in neural networks. We pass the input through the activation function before passing it to the next layer.

The net input value can be anything between `-inf` to `+inf`, and the neuron doesn't know how to bound the values, thus unable to decide the firing pattern. The activation function decides whether a neuron should be activated or not to bound the net input values.

Most common types of Activation Functions:

- Step Function
- Sigmoid Function
- ReLU
- Leaky ReLU

How would you build a restaurant recommendation on Facebook?

The answer is completely up to you. But before answering, you need to consider what business goal you want to achieve to set a performance metric and how you are going to acquire the data.

In a typical machine learning system design, we:

- Collect, clean, and analyze the data.
- Perform feature engineering
- Select a methodology, algorithm, or machine learning model
- Train and evaluate the performance on test and validation datasets.
- Streamline the processes and deploy the model in production.

You need to make sure you are focusing on design rather than theory or model architecture. Make sure to talk about model inference and how improving it will increase the overall revenues.

Also, give an overview of why you selected a certain methodology over the other.

Learn more about [building recommendation systems](#) by taking a course on DataCamp.

Given two strings A and B, write a function can_shift to return whether or not, A can be shifted some number of places to get B.

Solving [coding challenges](#) and working on your Python skills will improve your chance of getting past the coding interview stage.

Before jumping into solving a problem, you need to understand the question. You simply need to create a boolean function that will return True if by shifting the alphabets in String B, you get String A.

```
A = 'abid'
B = 'bida'
can_shift(A, B) == True
```

POWERED BY  datacamp

- Return false if the length of the string is not similar.
- Loop around the range of length of String A
- Create mut_a to create various combinations of characters using the String A
- During the loop, if mut_a is equal to String B returns True, else returns false.

```
def can_shift(a, b):
    if len(a) != len(b):
        return False
    for i in range(len(a)):
        mut_a = a[i:] + a[:i]
        if mut_a == b:
            return True
    return False
```

```
A = 'abid'
B = 'bida'
print(can_shift(A, B))
>>> True
```

POWERED BY  datacamp

What is Ensemble learning?

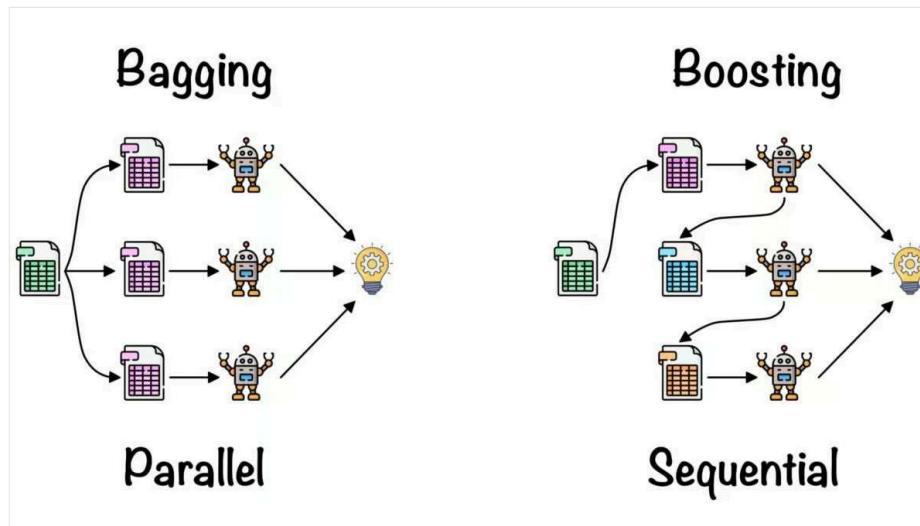
Ensemble learning is used to combine the insights of multiple machine learning models to improve the accuracy and performance metrics.

Simple ensemble methods:

- Mean/average: we average the predictions from multiple high-performing models.
- Weighted average: we assign different weights to machine learning models based on the performance and then combine them.

Advance ensemble methods:

- Bagging is used to minimize variance errors. It randomly creates the subset of training data and trains it on the models. The combination of models reduces the variance and makes it more reliable compared to a single model.
- Boosting is used to reduce bias errors and produce superior predictive models. It is an iterative ensemble technique that adjusts the weights based on the last classification. Boosting algorithms give more weight to observations that the previous model predicted inaccurately.



Bagging and Boosting by [Fernando López](#)

Learn more about averaging, bagging, stacking, and boosting by completing the [Ensemble Methods in Python](#) course.

Conclusion

As we conclude our exploration of essential machine learning interview questions, it's evident that succeeding in such interviews requires a blend of theoretical knowledge, practical skills, and an awareness of the latest trends and technologies in the field. From understanding the basic concepts like semi-supervised learning and algorithm selection, to delving into the complexities of specific algorithms like KNN, and grappling with role-specific challenges in NLP, computer vision, or reinforcement learning, the scope is vast.

Whether you are a beginner looking to break into the field or an experienced practitioner aiming to advance further, continuous learning and practice are key. DataCamp offers a comprehensive [Machine Learning Scientist with Python](#) track that provides a structured and in-depth way to enhance your skills.

Earn a Top AI Certification

Demonstrate you can effectively and responsibly use AI.

Get Certified, Get Hired

FAQs

What is a machine learning interview like?

Typically, machine learning technical interviews are divided into multiple parts:

1. Coding interview
2. Algorithms and Data processing
3. Role-specific interview
4. ML system design interview
5. Machine learning operations and best practices

The non-technical or on-site interviews are also part of the machine learning interview process, but they are more general and company-specific.

Is it easy to pass machine learning interviews at Amazon?

Do machine learning engineers have coding interviews?

What are the main tasks performed by an AI/ML engineer?

How do I prepare for a machine learning interview?

TOPICS

Machine Learning Data Science



Training more people?

Get your team access to the full DataCamp for business platform.

For Business

For a bespoke solution [book a demo](#).

Machine Learning Courses

COURSE

Machine Learning with PySpark

4 hr 23.7K

Learn how to make predictions from data with Apache Spark, using decision trees, logistic regression, linear regression, ensembles, and pipelines.

[See Details →](#)[Start Course](#)[See More →](#)

Related

**BLOG**[Top 30 Cloud Computing Interview Questions and...](#)**BLOG**[Top 30 AI Interview Questions and Answers For All Skill Levels](#)**BLOG**[Top 30 SQL Server Interview Questions \(2025\)](#)[See More →](#)

Grow your data skills with DataCamp for Mobile

Make progress on the go with our mobile courses and daily 5-minute coding challenges.



LEARN

[Learn Python](#)[Learn AI](#)[Learn Power BI](#)[Learn Data Engineering](#)[Assessments](#)[Career Tracks](#)[Skill Tracks](#)[Courses](#)[Data Science Roadmap](#)

DATA COURSES

[Python Courses](#)[R Courses](#)[SQL Courses](#)

[Power BI Courses](#)[Tableau Courses](#)[Alteryx Courses](#)[Azure Courses](#)[AWS Courses](#)[Google Sheets Courses](#)[Excel Courses](#)[AI Courses](#)[Data Analysis Courses](#)[Data Visualization Courses](#)[Machine Learning Courses](#)[Data Engineering Courses](#)[Probability & Statistics Courses](#)

DATALAB

[Get Started](#)[Pricing](#)[Security](#)[Documentation](#)

CERTIFICATION

[Certifications](#)[Data Scientist](#)[Data Analyst](#)[Data Engineer](#)[SQL Associate](#)[Power BI Data Analyst](#)[Tableau Certified Data Analyst](#)[Azure Fundamentals](#)[AI Fundamentals](#)

RESOURCES

[Resource Center](#)[Upcoming Events](#)[Blog](#)[Code-Alongs](#)[Tutorials](#)

[Docs](#)[Open Source](#)[RDocumentation](#)[Book a Demo with DataCamp for Business](#)[Data Portfolio](#)

PLANS

[Pricing](#)[For Students](#)[For Business](#)[For Universities](#)[Discounts, Promos & Sales](#)[DataCamp Donates](#)

FOR BUSINESS

[Business Pricing](#)[Teams Plan](#)[Data & AI Unlimited Plan](#)[Customer Stories](#)[Partner Program](#)

ABOUT

[About Us](#)[Learner Stories](#)[Careers](#)[Become an Instructor](#)[Press](#)[Leadership](#)[Contact Us](#)[DataCamp Español](#)[DataCamp Português](#)[DataCamp Deutsch](#)[DataCamp Français](#)

SUPPORT

[Help Center](#)[Become an Affiliate](#)



[Privacy Policy](#) [Cookie Notice](#) [Do Not Sell My Personal Information](#) [Accessibility](#) [Security](#) [Terms of Use](#)

© 2025 DataCamp, Inc. All Rights Reserved.