

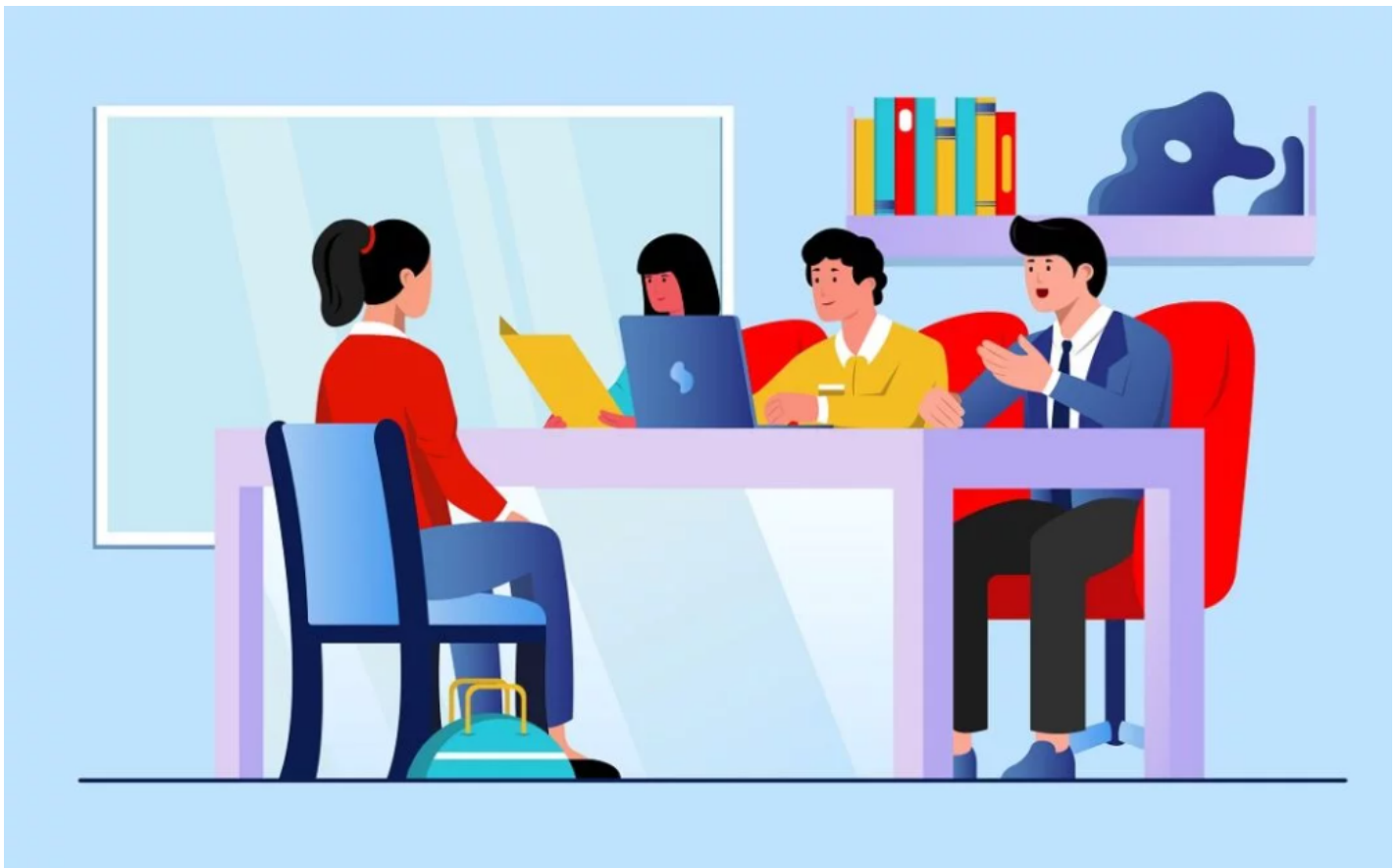
Interview Questions on KNN in Machine Learning

[ALGORITHM](#)[INTERMEDIATE](#)[INTERVIEW QUESTIONS](#)[INTERVIEWS](#)[MACHINE LEARNING](#)

This article was published as a part of the [Data Science Blogathon](#).

Introduction

K nearest neighbors are one of the most popular and best-performing algorithms in supervised machine learning. Furthermore, the KNN algorithm is the most widely used algorithm among all the other algorithms developed due to its speed and accurate results. Therefore, the data science interviews might ask in-depth questions about the k nearest neighbors. In this article, we will discuss and solve advanced interview questions related to the [k nearest neighbors](#) in machine learning.

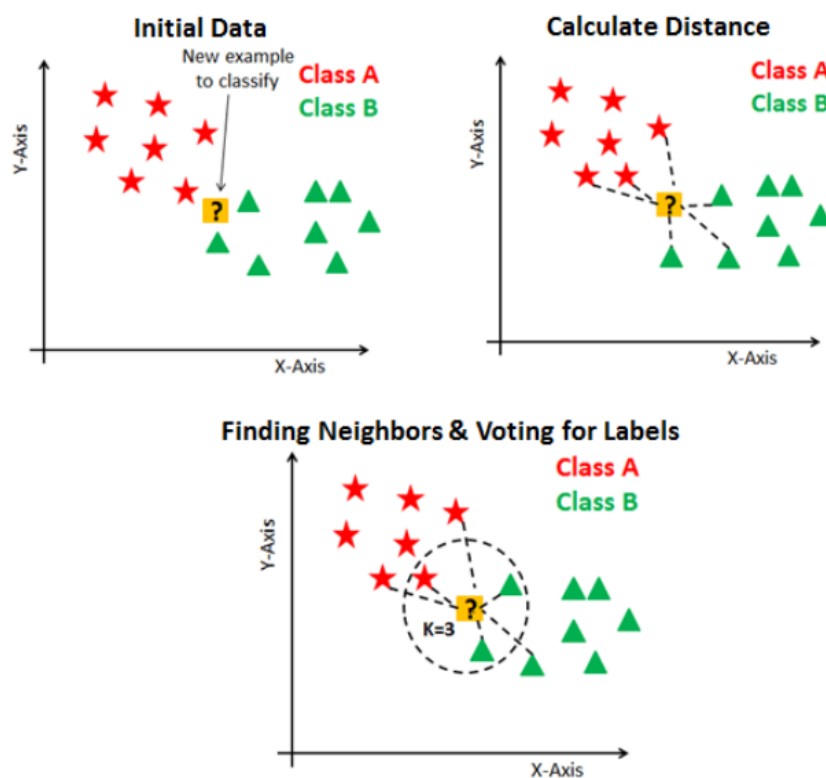


Source: <https://resources.biginterview.com/wp-content/uploads/2022/07/Panel-Interview-101-1080x675.jpg>

1. Why is the Time Complexity Very High For the Prediction Phase in KNN? Explain with Reasons.

In almost every machine learning algorithm, the algorithm trains first on the training data and then makes predictions based on the dataset it was prepared before. K nearest neighbor is a machine learning clustering algorithm that divides the training data into a particular number of clusters by calculating the distance of the specific points from other points. Then while predicting for careful observation, it again calculates the length of the issue and tries to settle the matter in a particular cluster to make predictions.

There are two machine learning algorithms: **Lazy Learning** and **Eager Learning**. Lazy learning is a machine learning algorithm that does not train on the training data provided. Instead, when the query is made to the algorithm to predict for it, it only trains on the training dataset. While in eager learning algorithms, the algorithm tries to teach the training data when provided. Then, when the new query is made for prediction, the algorithm predicts based on the training on the previous data. K nearest neighbor also stores the training data. Then, when there is a time for the prediction phase, this algorithm calculates the distances of the query point from other points and tries to assign the cluster to the particular topic. So it only trains on the data when a query is made to the system, which is why it is known as a lazy learning algorithm.



Source: <https://qph.cf2.quoracdn.net/main-qimg-53d74e4e12547a448799d5ebb126ebfc>

- As lazy learning algorithms store the data, it requires more space. That is why KNN requires more space to store.
- The reason behind the speed of the KNN is that it does not train on the training data, so training in KNN is very fast.
- As KNN trains on the training data while the prediction phase, predictions tend to be very slow in the KNN algorithm.

2. Why is KNN Algorithm Said to be More Flexible?

The K's nearest neighbor is the non-parametric algorithm, which does not make any primary assumption while training and testing on the dataset. The parametric machine learning algorithms like linear regression, logistic regression, and naive Bayes make primary assumptions like the data should be linear or there should be no multicollinearity in the dataset. Due to this, some of the algorithms could be more flexible. If the assumptions they made are satisfied, we can only use them. For example, if the data is not linear, then linear regression cannot be applied; if the dataset has multicollinearity, then naive Bayes can not be applied.

But in the case of the KNN algorithm, as it is a non-parametric algorithm, it does not make any primary assumption on the dataset so that it can apply it to any dataset, and also it returns good results. So this is the main reason behind the flexibility of the KNN algorithm.

3. Why is KNN Algorithm Less Efficient Than Other Machine Learning Algorithms?

If you prefer flexibility, then KNN would be the best fit for the problem statement, but it also has a drawback in efficiency. Suppose one wants efficiency for the particular model. In that case, one should go for other algorithms available as KNN is not a very efficient machine learning algorithm compared to the different machine learning algorithms. As KNN is a lazy learning algorithm, it generally stores the input or the training data and does not train while the training data is fed.

Instead, it trains when the query for prediction is made, which is the main reason behind the more time complexity in the prediction phase. While some of the eager learning algorithms, like linear regression, instantly train on the training data and predict the data very fast. So that due to this reason, KNN is said to be less efficient compared to the other machine learning algorithms.

4. Why Does KNN Performs Well on Normalized Datasets?

We know that K's nearest neighbor is the distance-based machine learning algorithm, which calculated the euclidian distance between points and returned the output. Noe, in some cases, the scale of the features of the dataset might be very different; in that case, the distances between points will also be very high or very low. Due to this, there will be errors or noisy data in euclidian lengths; hence, the algorithm will not perform well. For example, we have a dataset of the Age and Salary of the person, now, the Age may vary from 0 to 99, and the salary can be in lakhs or crores. So here, the scale is very different between the two features, so it also affects the euclidian distance, and hence the algorithm will perform poorly if the data is not normalized. Now, if the data is normalized, in that case, all the values will be between 0 and 1. So calculating the euclidian distances on the same scale of the data will be very easy o=for the algorithm, and hence the model will perform well.

5. How Could the Less Value of K Lead to Overfitting in the KNN Algorithm? Explain.

The value of K in the KNN algorithm means the number of neighbors. So suppose the value of k is 3. Then we want to consider three neighbors for the model's training. Let's have a case where we have a significantly less value of k, say 1. We will only consider one neighbor for training the model in this case. So many clusters will be created, and the model will try to fit every data point of the dataset, leading to good performance on training data and poor performance on testing data.

On the other side, if we have a very high value of K , then there will be less amount of clusters that will be created, which will lead to the abysmal performance of the model; the case of under-fitting, where the model will perform poorly on training and testing both data.

Conclusion

In this article, we discussed advanced interview questions related to the k nearest neighbors and their solutions with core intuitions and logical reasons behind them. Knowledge about these concepts will help one answer these tricky and different questions efficiently.

Some **Key Takeaways** from this article are:

1. KNN is a lazy learning algorithm that stores the data while the training phase and does not use the stored data while the training phase. While in the predictions phase of KNN, so many calculations are involved as it is a lazy learning algorithm.
2. The time complexity for KNN in the training phase is low, and the testing phase is high, as it is a lazy learning algorithm which never does any calculations while the training phase. The space complexity also follows the same trend as time complexity in the KNN algorithm.
3. KNN is a non-parametric machine learning algorithm that provides higher flexibility and lower efficiency. As it is a non-parametric algorithm, it has no pre-assumptions like linear regression.

The media shown in this article is not owned by Analytics Vidhya and is used at the Author's discretion.

Article Url - <https://www.analyticsvidhya.com/blog/2022/12/interview-questions-on-knn-in-machine-learning/>

[Parth Shukla](#)



UG (PE) @PDEU | 25+ Published Articles on Data Science | Data Science Intern & Freelancer | Amazon
ML Summer School '22 | AI/ML/DL Enthusiast | Reach Out @portfolio.parthshukla.live