

## Dive Deep into KNN

1. Explain k-NN intuitively?

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2927/k-nearest-neighbours-geometric-intuition-with-a-toy-example/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

2. Implement k-NN in Python/Pseudo-code (simple test for your programming knowledge)

Refer:

<https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>

3. Give examples of cases where kNN would not perform as well.

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2928/failure-cases-of-knn/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

4. Define Manhattan distance, and where it is preferred over euclidean distance.

HINT: outliers

5. How are Manhattan and Euclidean distance related to Minkowski distance?

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2929/distance-measures-euclideanl2-manhattanl1-minkowski-hamming/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

6. Where is the Hamming distance preferred in the real world?

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2929/distance-measures-euclideanl2-manhattanl1-minkowski-hamming/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

7. Cosine similarity: intuition, math and where is it used?

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2930/cosine-distance-cosine-similarity/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

8. Derive the relationship between cosine and euclidean distance.

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2930/cosine-distance-cosine-similarity/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

9. Why do we need test set?

Refer:

<https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/293>

[1/how-to-measure-the-effectiveness-of-k-nn/3/module-3-foundations-of-natural-language-processing-and-machine-learning](https://www.appliedaigcourse.com/lecture/11/applied-machine-learning-online-course/2932/testevaluation-time-and-space-complexity/3/module-3-foundations-of-natural-language-processing-and-machine-learning)

10. Time and Space complexity of kNN at train time and evaluation time?

Refer:

<https://www.appliedaigcourse.com/lecture/11/applied-machine-learning-online-course/2932/testevaluation-time-and-space-complexity/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

11. When not to use kNN in the real world?

Refer:

<https://www.appliedaigcourse.com/lecture/11/applied-machine-learning-online-course/2933/knn-limitations/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

12. What happens to decision surface as k changes?

Refer:

<https://www.appliedaigcourse.com/lecture/11/applied-machine-learning-online-course/2934/decision-surface-for-k-nn-as-k-changes/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

13. How to find the optimal K in kNN.

HINT: Cross validation

14. Graphs between K vs error/loss.

15. When do we NOT use random split to create train, CV and test data?

Refer:

<https://www.appliedaigcourse.com/lecture/11/applied-machine-learning-online-course/2940/time-based-splitting/3/module-3-foundations-of-natural-language-processing-and-machine-learning>

16. How to obtain optimal weights in weighted-kNN

17. Why do we need to use a kd-Tree?

18. Explain how kd-tree works intuitively?

19. Time and Space complexity of kd-tree.

20. When not to use a kd-tree?

21. Give an intuitive explanation of LSH? Where should we use LSH over brute-force and kd-tree?

22. How to perform multi-class classification using kNN?

23. Does data imbalance impact kNN? If so, how to fix it?

24. How to find outliers using kNN? {some interviewers may not know LOF}

25. How to use kNN for imputing missing values?

26. How does kNN work if the dimensionality of the data is large like in text data?

27. Why not use forward feature selection while computing feature importance?

28. Explain bias variance tradeoff intuitively?

More references:

[https://medium.com/@cornell\\_data/interview-case-study-1-sampling-methods-and-parameter-changes-4799c580aa42](https://medium.com/@cornell_data/interview-case-study-1-sampling-methods-and-parameter-changes-4799c580aa42)

[https://medium.com/@cornell\\_data/interview-case-study-2-no-free-lunch-b0b9d1e6dbd2](https://medium.com/@cornell_data/interview-case-study-2-no-free-lunch-b0b9d1e6dbd2)