**Dive Deep into KNN**

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

[Develop k-Nearest Neighbors in Python From Scratch (machinelearningmastery.com)](https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/)

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

HINT: outliers

D) Where is the Hamming distance preferred in the real world?

E) Time and Space complexity of kNN at train time and evaluation time?

F) When not to use kNN in the real world?

G) . How to find the optimal K in kNN.

HINT: Cross validation

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

I) Why do we need to use a kd-Tree?

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

K) Does data imbalance impact kNN? If so, how to fix it?

L) How does kNN work if the dimensionality of the data is large like in text data?

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

N) Graphs between K vs error/loss.

O) Cosine similarity: intuition, math and where is it used?

P) Why do we need test set?

**Interview Questions :**

   
In k-means or kNN, we use euclidean distance to calculate the distance between nearest neighbours. Why not manhattan distance ?(**https://www.analyticsvidhya.com/blog/2017/09/30-questions-test-k-nearest-neighbors-algorithm/**)

 How to test and know whether or not we have overfitting

 How is kNN different from k-means clustering?(**https://stats.stackexchange.com/questions/56500/what-are-the-main-differences-between-k-means-and-k-nearest-neighbours**)

 Can you explain the difference between a Test Set and a Validation Set?(**https://stackoverflow.com/questions/2976452/whats-is-the-difference-between-train-validation-and-test-set-in-neural-netwo**)

 How can you avoid overfitting in KNN?

**More references:**

[https://medium.com/@cornell\_data/interview-case-study-1-sampling-methods-and-para meter-changes-4799c580aa42](https://medium.com/@cornell_data/interview-case-study-1-sampling-methods-and-para%20meter-changes-4799c580aa42)

<https://medium.com/@cornell_data/interview-case-study-2-no-free-lunch-b0b9d1e6dbd2>