29.10 K-NN Limitations

a) **High Space Complexity**
   Space Complexity of KNN = O(nd)
   Where 'n' → number of data points, 'd' → number of dimensions
   Let us now consider the Amazon Fine Food Reviews Dataset where n = 364K and d = 100K (let us assume, because we use techniques like BOW/TF-IDF)
   Now the data matrix contains 364K * 100K elements.
   So the total space needed = 364K * 100K = ~36GB of RAM is required. So the space complexity is so high.

b) **High Time Complexity**
   In Amazon Fine Food Reviews Dataset, if a review is given, the system should be able to give the result within less than a millisecond.
   But here the system has to perform 36 billion computations. So the time complexity is so high.

Solutions to address this problem

a) The simple implementation of KNN has a time complexity of O(nd) and a space complexity of O(nd). So we have to implement KNN with better time and space complexities. **(or)**
b) Just change the algorithm. KNN is simply intuitive and elegant, but the reason for not being used is because of large time and space complexities.

Note: When we process such a huge amount of data on distributed systems, we'll have multiple threads processing the data. But still the time complexity can't be affected by the usage of multiple threads, whereas the total computation time reduces.