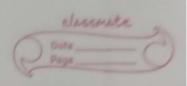


Ouestion 4 Algorithm submitted. Eroof for time complexity of O (nd) fer testing hoint O(n*d)Use quikeelest to find kth emallest distonce O(n)Einding all nearest reighbours to store their

y-values: O(n)Since nd is more significant compared to n, being dominant, total time complexity = O(n'd) viruspetus

of k 2. When does each afferough perform letter, and why! The least squares linear regression is letter when the data to be predict continuous numeric data where there is a linear relationship between independent and dependent variables Eg. Data in dataset D The k-Nearest Neighbour can be used for classification as well as negression. For negression, it doesn't assume bright and as such is very suitable for complex non linear dataset. However, computational cost becomes excessive for large datasets Eg. Pataset E.



Which afferoach is better and why?

The k-nearest neighbours for form better as they have more it a total higher k value. As the k value increase there are more near neighbour using the neighbour with y value more than only using the neighbour with the lowest le distance. The linear regression prediction is better than k= or 2 leut romains the M SE remains constant as the k value go will down.

Irefecting the rearest distances between the training and test set data, a large number of nearest neighbours are quite close to the data point tranging from 2.8 To 32 locause of which this approach is better than linear regression for this set of data points with d=20