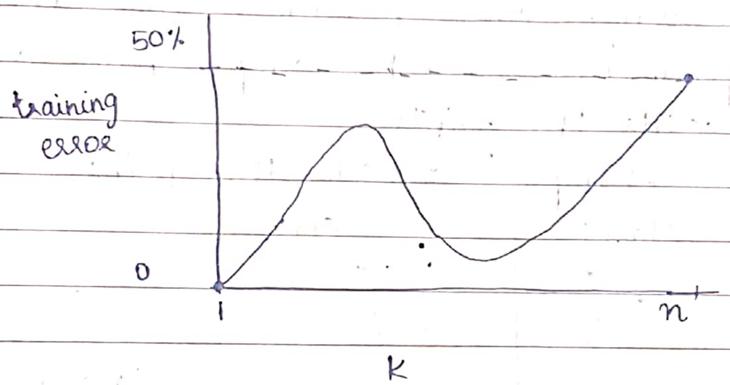
DATE □□□□□□□□

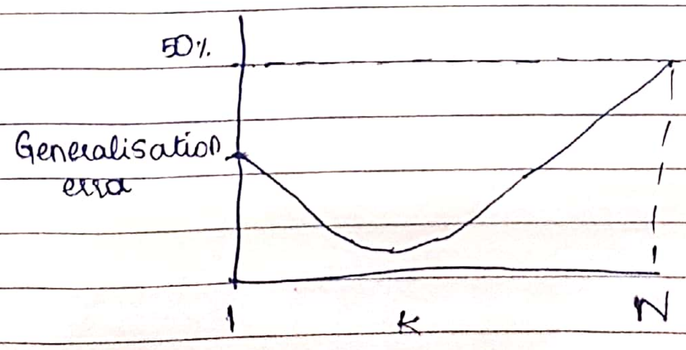
1. a) Assuming a evenly split model where the overlapped region consists of points from both classes in the same count (Not assuming in the overlap of class 1 Side, points of class 1 is more or something like that)

When we take k='n', the accuracy as well as training error is 50%. As 'k' reduces, we begin to take less points from the other class so accuracy increases (aka) training error reduces. This stage continues until the point where we reach the overlapped region where training error increases or accuracy decreases and when k=1, the training error becomes 0 as each point would ~~becom~~ vote for itself.



The dip and the slope of the curve depends on the Overlapping region

b) Having the same assumption as above, we get the generalisation error ~~increases~~ reduces upto a certain value as k moves from 1 but as we get closer to the overlap region, the error increases and achieves the least efficiency of 50% at k=n. This is be cause as we get closer to the overlap region the error increases which in-turn affects the overall accuracy from being a ~~perfect 100~~ big in case of K=~~N~~1 ~~[illegible]~~.



The dip and slope of the curve depends on the points and the Overlapping region

classmate

PAGE □□□

DATE

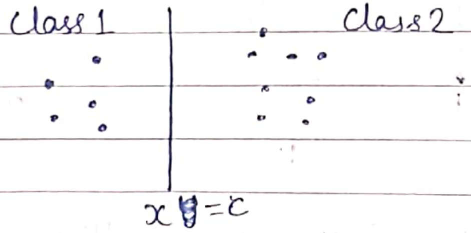
(c) when the number of dimensions is high, it fails because of

(i) at a higher $\overset{\text{dimension}}{\cancel{\text{distance}}}$, distance formulas dont work and all the points are concentrated on the corners of the graph instead of being spread across the plane

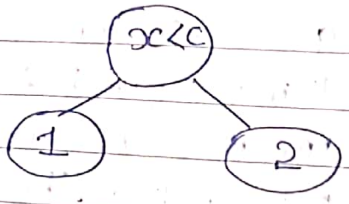
(ii) As the number of dimensions increase, the $\overset{\text{computation ti}}{\cancel{\text{time complex}}}$ increases i.e. the time taken to compute the distance with each point also tends to increase. i.e. it is very computationally intensive

d) Yes it is possible to build a decision tree for 1-nn. Assume the case where the points are separated by a straight line. In such a case, the decision tree parameters could be simply based on the line that splits the points.

example: consider the points



The decision tree could be:



classmate

Scanned by CamScanner