**Report**

Name : Shashwat Pankajkumar Chauhan

Student id Number : 1002034494

Net id : spc4494

Question\_1\_(C) : **Compare the error results and try to determine for what “function depths” overfitting might be a problem. Which ”function depth” would you consider the best prediction function and why? For which values of k and d do you get minimum error?**

Answer : The minimum error that I am getting is k = 1 and d=6 (depth) , Error : 0.49619527464385926 .Thus the minimum error gets overfitting in this dataset . From k = 1 and depth 6 we are getting the minimum error .

Text, letter

Description automatically generated

Question\_1\_(D) : **Repeat the experiment and evaluation of part b) and c) using only the first 20 elements of the training data set part b) and the Test set of part c). What differences do you see and why might they occur?**

Answer : If we compare the error of question A and B because A has 128 dataset as well as B has trained with 20 dataset . So Error we are getting in 20 dataset is more compared to 120 dataset . In 20 dataset the randomness is higher compared to 128 .

Graphical user interface, text, application

Description automatically generated with medium confidence

Question 2 :

Question\_2\_(C) : **How does the performance compare to the one for the results from Question 1 (C).**

Answer : This function performs better because it takes weightaed value of point other from from 1\_c .

**Text

Description automatically generated**

Question\_2\_(D) : **How does the performance compare to the one for the results from Question 1 d) ? Why might this be the case?**

Answer : If we compare the error of question A and B because A has 128 dataset as well as B has trained with 20 dataset . So Error we are getting in 20 dataset is more compared to 120 dataset . In 20 dataset the randomness is higher compared to 128.

Graphical user interface, text

Description automatically generated

Question\_2\_(E) :

Answer : The results from parts c) and d) , it is evident that the dataset used is not derived from a function that is consistant with the function format in question 1. This is because of the significant increase in the error.

Question 3 :

Question\_3\_(C) : **Discuss what differences exist and why one method might out perform the others for this problem.**

Answer : Logestic regression performs better because it cames more values, thus the model get trained for more wider range and make it accurate .

A picture containing background pattern

Description automatically generated

Question\_3\_(D) : **Again, discuss what differences exist and why one method might outpe-rform the others in this case.**

Answer : Yes, removing age makes it better because age parameter adds randomess into ago because can’t be factor to determine the gender

Chart, scatter chart

Description automatically generated