

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 .

```
/Q1_C.py  
START Q1_C
```

```
[[1, 0, 0.45570554646571804], [1, 1, 0.45385138581055556], [1, 2, 0.454192467465364], [1, 3, 0.4570189664123979], [1, 4, 0.446390059397  
7806], [1, 5, 0.44146504240746137], [1, 6, 0.4401322266994109], [2, 0, 0.45570554646571804], [2, 1, 0.455411382523839], [2, 2, 0.445462  
63639727185], [2, 3, 0.4452111322931112], [2, 4, 0.44634608176185486], [2, 5, 0.4463588064176677], [2, 6, 0.44592498059980773], [3, 0,  
0.45570554646571804], [3, 1, 0.45829567617207395], [3, 2, 0.4582686545191603], [3, 3, 0.4585013328136981], [3, 4, 0.45835582053024454],  
[3, 5, 0.4583505100443971], [3, 6, 0.4585845086079231], [4, 0, 0.45570554646571804], [4, 1, 0.4471240785274213], [4, 2, 0.448851222176  
6477], [4, 3, 0.4487511531583508], [4, 4, 0.4487637228139002], [4, 5, 0.4487232449575752], [4, 6, 0.4487295497473802], [5, 0, 0.4557055  
4646571804], [5, 1, 0.4570737322384298], [5, 2, 0.4557011647709517], [5, 3, 0.45584365802138266], [5, 4, 0.45660547894631787], [5, 5, 0  
.45370119092773975], [5, 6, 0.4562215464827606], [6, 0, 0.45570554646571804], [6, 1, 0.45594741732981664], [6, 2, 0.4558037357879302],  
[6, 3, 0.45628739284234626], [6, 4, 0.4563469061532261], [6, 5, 0.4593786427200399], [6, 6, 0.4547313545316895], [7, 0, 0.4557055464657  
1804], [7, 1, 0.45570904344044216], [7, 2, 0.45612523377325126], [7, 3, 0.4569035648899327], [7, 4, 0.4569098868844659], [7, 5, 0.45695  
811156662686], [7, 6, 0.4568293485419493], [8, 0, 0.45570554646571804], [8, 1, 0.455663607568168], [8, 2, 0.45566650903554656], [8, 3,  
0.4556653920894691], [8, 4, 0.45340674619392224], [8, 5, 0.45330378572322383], [8, 6, 0.45220133722588257], [9, 0, 0.45570554646571804]  
, [9, 1, 0.45526639715625034], [9, 2, 0.4557364191168053], [9, 3, 0.4554115835155138], [9, 4, 0.4509524579894433], [9, 5, 0.45106255280  
77436], [9, 6, 0.4522588642610488], [10, 0, 0.45570554646571804], [10, 1, 0.4558285085633528], [10, 2, 0.45583318945541157], [10, 3, 0.  
4582426332990076], [10, 4, 0.45791656217294063], [10, 5, 0.45789438845499364], [10, 6, 0.4579551822166078]]  
least error was:- 0.22006611334970544  
END Q1_C
```

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 .

```
for k=10, d=3, error=0.6480396860015915  
for k=10, d=4, error=0.6503268699897489  
for k=10, d=5, error=0.5491737223925609  
for k=10, d=6, error=0.5428655496031918  
  
least error was:- 0.5428655496031918  
END Q1_D
```

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 weighted value of point other from from 1_c .

```
START Q2_C  
the error: 0.2172236058992584  
END Q2_C
```

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.

```
START Q2_D  
the error: 0.4720431347903462  
END Q2_D  
PS C:\Users\shash\OneDrive\Desktop\ML Project 2\netId_project_2>
```

Question 3 :

Question_3_(C) : Discuss what differences exist and why one method might out perform the others for this problem.

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

```
401], [0.9327731092436975], [0.0005322861152998303, -0.8143907190829495, 3.75249975445926], [0.5630252100840336], [[-0.07989589227340649, -1.0159878556130701, 4.45581
2294070773], [0.5630252100840336], [[-0.038061483990235906, -1.4686830422841876, 3.911247963180987], [0.9327731092436975], [[-0.06866668497790839, -1.427640492176132,
3.8414854245861254], [0.907563025210084], [[-0.03462800294851692, -0.9326907320393109, 4.170661290044754], [0.5714285714285714], [0.05113836031774886, -1.58381715685
3.8414854245861254], [0.907563025210084], [[-0.03462800294851692, -0.9326907320393109, 4.170661290044754], [0.5714285714285714], [0.05113836031774886, -1.58381715685
031210318, 4.123559966577222], [0.907563025210084], [[-0.0571501365229546, -0.8740635063213619, 4.448843885174777], [0.5462184873949579], [[-0.04833382807036536, -1.065981578454323, 4.9467604128140374], [0.5630252100840336], [[-0.0429671374484845
16, -1.5708977504544968, 4.162620383751559], [0.9327731092436975], [[-0.01453761608120788, -1.6246497572454408, 4.295057797919675], [0.9411764705882353], [[0.03901746408426342, -1.5998349850316544, 4.23380042886366], [0.9411764705882353], [[0.00
949987332976926, -0.8062292907880927, 3.552556468257101], [0.5630252100840336], [[0.002095476295175057, -1.276893608268788, 3.3652838379201255], [0.9411764705882353], [[0.022131902628426223, -0.8967930147564843, 4.574147453765891], [0.5462184873
949579], [[-0.05547831703765326, -0.8525957065970766, 4.290252816635339], [0.546218487394958], [[-0.0432686973352156, -0.8622171433000937, 2.5182734400720546], [0.8235294117647058], [[0.08839525361397528, -0.8899605751279074, 3.9400057812224776
], [0.5714285714285714], [[0.11524359762406153, -1.3146971339212485, 4.117035128958795], [0.7394957983193278], [[-0.021124314694325057, -1.2354815698631962, 3.980587257077749], [0.680672268907563], [[-0.037906748540738305, -1.5540257389178, 4.16
7971439169996], [0.907563025210084], [[-0.033494946344008364, -1.7469669379224286, 4.623301526914966], [0.9327731092436975], [[-0.026244624868864228, -0.8911427681273642, 4.517964449601211], [0.546218487394958], [[-0.04821247539622865, -1.6378618
844919646, 4.33192670971144], [0.9411764705882353], [[-0.026976522871565153, -1.593470805580699, 4.315301059117578], [0.891596638655462], [[-0.062084373958341796, -0.793287422259005, 4.215491955621813], [0.5462184873949579], [[-0.1058361903400
1432, -1.46647948234533, 3.921079258541023], [0.9243697478991597], [[-0.0971879244381863, -1.603975070175747, 4.2693087265322465], [0.9327731092436975], [[-0.041283653350363485, -0.6517526963746818, 3.884808289525272], [0.5462184873949579], [[-
0.0398081338491052, -0.8797381765248272, 4.538692790021668], [0.5546218487394958], [[-0.03916672168378696, -1.5118459899334606, 4.004306086512951], [0.9327731092436975], [[-0.09174270184346223, -1.6357579352672862, 4.3754181890118105], [0.9243697
478991597], [[-0.0362426868268554, -1.6372626002095064, 4.352062882398445], [0.9327731092436975], [[0.009644462895606195, -1.0434895452334185, 3.318672699470857], [0.6974789915966386], [[-0.07246246248396387281, -1.5499410667380604, 4.136922394731
162], [0.9327731092436975], [[-0.12111630693542591, -1.092923852087707, 3.4639633818662756], [0.7226890756302521], [[-0.03564218533906828, -1.1697368741280703, 3.6097904716028855], [0.7647058823529411], [[-0.06302147996301186, -1.319680448795087,
3.654877575643956], [0.8991596638655462], [[-0.059396665058124584, -0.8741864945278447, 4.470522604600225], [0.5462184873949579], [[-0.03810471952795863, -1.2679472854494605, 3.90103659126165], [0.7563025210084033], [[-0.03920865900826394, -0.74
40006806984706, 3.9678618111629382], [0.5546218487394958], [[0.0014416555157007443, -0.6480690278967697, 3.6760754925418637], [0.5462184873949579], [[-0.061672998132016286, -0.751396155349001, 3.9732465860898327], [0.5546218487394958], [[-0.02734
489274644247, -1.530493077644095, 4.05312642958711], [0.9327731092436975], [[-0.0594727582540199, -0.8988565886261313, 4.4479370059002572], [0.5462184873949579], [[-0.06416610031272252, -1.616457243690384, 4.289632144682793], [0.9411764705882353],
], [[-0.014086668510927546, -1.5891511081464603, 4.420101151178473], [0.8907563025210085], [[-0.04589717536301366, -1.5954825070260152, 4.21870691223877], [0.9411764705882353], [[-0.027495451132928655, -1.500082658578766, 4.2834318774816555], [0.
0571428571428571], [[-0.041709403619490326, -1.6107244075421172, 4.295053132008096], [0.9327731092436975], [[-0.061552941351368, -1.45608525337243, 3.917085835067477], [0.8991596638655462], [[-0.091514708009051787, -1.1051825876908952, 3.4443
25248904946], [0.7478991596638656], [[-0.01711973734083233768, -1.0825409395043324, 3.3000045307975823], [0.773109243697479], [[-0.0714487798769536, -1.4723311765281903, 3.9147932650374218], [0.9327731092436975], [[-0.07740966851017454, -1.5462545
308516478, 4.143950857975918], [0.9159663865546218], [[-0.0525002541976614, -0.7810228829259097, 3.946140862356827], [0.5462184873949579], [[-0.03759322330289214, -1.5680326881133735, 4.225854307593638], [0.8991596638655462], [[-0.06382633079539
687, -1.4416963964263405, 3.8322508742738166], [0.9327731092436975], [[-0.04162393632464245, -1.2293642566401686, 3.9927150019858], [0.680672268907563], [[-0.055198956656833956, -1.5021207913061736, 3.9882469060231966], [0.9327731092436975], [[-0.
04387859225251392, -1.579898385317558, 4.17789547880579], [0.9411764705882353], [[-0.034450809304069865, -1.5806772984995486, 4.243967452442124], [0.9159663865546218], [[-0.019668682026074465, -0.7256115784429914, 3.043869747140107], [0.5714285
714285714], [[-0.03419806106710442, -1.6120406137419555, 4.264324563601554], [0.9411764705882353], [[-0.026816603772607705, -0.7097002427199491, 4.450678036853995], [0.546218487394958], [[-0.022039583703829132, -2.147257044178415, 3.33002942287
34533], [1.0]]
END Q3_C
```

Question_3_(D) : Again, discuss what differences exist and why one method might outperform the others in this case.

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

