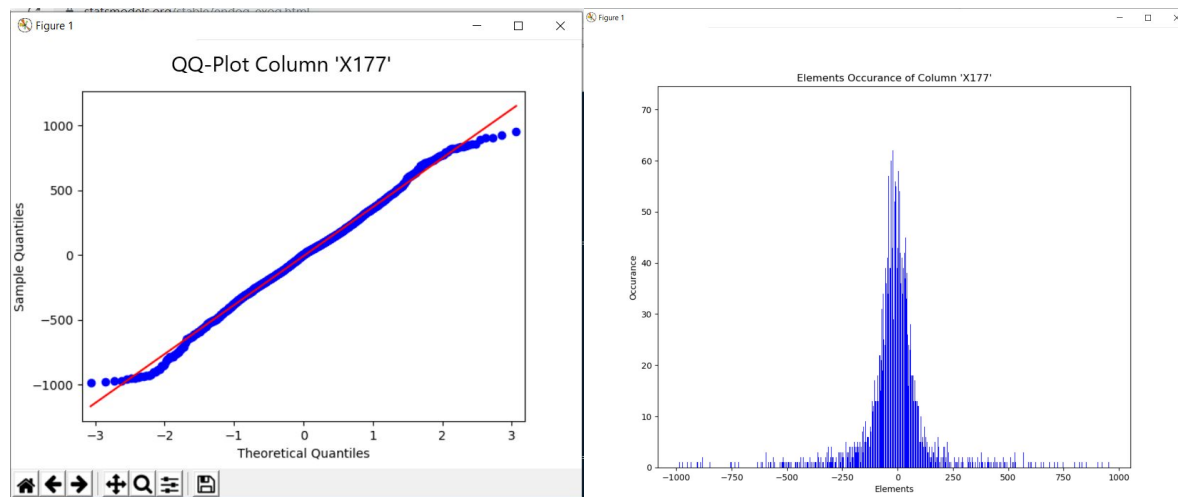


Jesse Hazard
Pawel Linek

Decision Tree Report

The problem we attempted to solve was how to create a decision tree classifier which can optimally predict epileptic seizures based on 178 features. We were challenged with doing this without using any direct library which implements a decision tree classifier(etc. sklearn).

Our implementation of this classifier includes creating a node class which holds the nodes name, an array of branches, and data. The node class also consists of a compare function. The node class is utilized by the larger Decision tree which is initialized with a data set. The Decision Tree contains the dataset, the attributes, and the head(root) node. There are also numerous functions part of the DecisionTree class including a traverse function which takes a column and traverses the tree until a classification [1-5] is found. There is also a classify function which calls the traverse function and attempts to find the previously mentioned classification. GrowTree function takes a dataset and a node and creates a decision tree that first the data. It “grows” the tree, and finally the majority function is used to calculate the output which appears most often. This is used when limiting the height of the tree in order to generalize much better. On top of these classes, we also have the get_attribute, info_gain, entropy, get_threshold, and remainder functions which are used to find the best column to split on, and which value in that column to split. It then returns the 2 partitions of the dataset which result when split on that threshold and attribute for repeated partitions of the data. Each partition is based on lowering the entropy and increasing the information gain of the resulting output column for the partitioned dataset. Some data pruning which we used included limiting the max height of the tree to 20, and due to the data in each column being a normal distribution as shown in the graphs below, we chopped off the bottom and top 25% of the data per column:



One major issue that arose toward the training portion of this assignment was that while training, we run out of features to split on if we trained on too many rows. This could be for several reasons, however, we weren't able to find the issue and correct it in the allotted time for this assignment. Therefore, when we trained our dataset, our train dataset could only consist of about 1000 columns. Due to this limitation, our classifier isn't the most accurate. Below is a sample of our confusion matrix when tested on 100 rows of our dataset which was our first attempt test:

	Guess1	Guess2	Guess3	Guess4	Guess5
0	568	557	422	26	56
1	1158	159	264	0	13
2	1148	149	283	0	14
3	956	245	364	1	74
4	1289	52	226	0	25

As can be seen, training on such a small dataset results in a low accuracy of about 13%. However, the model trained on 1000 data points performed better. Twice as well in fact, below is the 1000 datapoint confusion matrix:

	Guess1	Guess2	Guess3	Guess4	Guess5
0	728	225	272	238	166
1	218	175	493	314	394
2	197	153	478	360	406
3	277	246	401	338	378
4	148	162	527	302	453

The increase of data resulted in a 26% accurate decision tree classifier (which echoes the accuracy of kaggle). If we were able to fix our bug and train on the full dataset, I believe without a doubt that the classifier's accuracy would have increased much more.

We weren't able to get a decision tree diagram made for us, but we were able to print out our decision trees rules. The rest of the pages of this report are the rules for our 1000 datapoint trained decision tree:

RULES FOR DECISION TREE (Max Height 20)

$\&(X_{171} < -118.0) \&(X_{175} < -319.0) \&(X_1 < -362.5)$ then 1 with depth 4

$\&(X_{171} < -118.0) \&(X_{175} < -319.0) \&(X_1 > -362.5)$ then 1 with depth 4

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} < -230.0)$ then 1 with depth 4

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} < -147.0)$ then 1 with depth 5

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 < -129.0)$ then 1 with depth 6

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 < -136.0)$ then 1 with depth 7

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} < -12.4.0)$ then 1 with depth 8

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} > -12.4.0) \&(X_{57} < -99.5)$ then 1 with depth 9

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} > -12.4.0) \&(X_{57} > -99.5) \&(X_{11} < -64.5)$ then 4 with depth 10

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} > -12.4.0) \&(X_{57} > -99.5) \&(X_{11} > -64.5) \&(X_{12} < -62.5)$ then 3 with depth 11

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} > -12.4.0) \&(X_{57} > -99.5) \&(X_{11} > -64.5) \&(X_{12} > -62.5) \&(X_9 < -67.5)$ then 1 with depth 12

$\&(X_{171} < -118.0) \&(X_{175} > -319.0) \&(X_{177} > -230.0) \&(X_{39} > -147.0) \&(X_2 > -129.0) \&(X_8 > -136.0) \&(X_{10} > -12.4.0) \&(X_{57} > -99.5) \&(X_{11} > -64.5) \&(X_{12} > -62.5) \&(X_9 > -67.5)$ then 1 with depth 12

$\&(X_{171} > -118.0) \&(X_{163} < -98.5) \&(X_{178} < -99.5) \&(X_{89} < -186.0)$ then 1 with depth 5

$\&(X_{171} > -118.0) \&(X_{163} < -98.5) \&(X_{178} < -99.5) \&(X_{89} > -186.0) \&(X_{48} < -200.5)$ then 3 with depth 6

$\&(X_{171} > -118.0) \&(X_{163} < -98.5) \&(X_{178} < -99.5) \&(X_{89} > -186.0) \&(X_{48} > -200.5)$ then 1 with depth 6

$\&(X_{171} > -118.0) \&(X_{163} < -98.5) \&(X_{178} > -99.5) \&(X_{176} < -88.5) \&(X_3 < -86.0)$ then 2 with depth 6

$\&(X_{171} > -118.0) \&(X_{163} < -98.5) \&(X_{178} > -99.5) \&(X_{176} < -88.5) \&(X_3 > -86.0) \&(X_{110} < -64.5)$ then 3 with depth 7

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176<-88.5)&(X3>-86.0)&(X110>-64.5)&(X82<-78.5
) then 4 with depth 8

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176<-88.5)&(X3>-86.0)&(X110>-64.5)&(X82>-78.5
) then 4 with depth 8

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101<-68.5)&(X45<-92.5) then 3 with
depth 7

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101<-68.5)&(X45>-92.5)&(X63<-98.
5) then 5 with depth 8

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101<-68.5)&(X45>-92.5)&(X63>-98.
5)&(X105<-89.5) then 3 with depth 9

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101<-68.5)&(X45>-92.5)&(X63>-98.
5)&(X105>-89.5)&(X61<-88.0) then 3 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101<-68.5)&(X45>-92.5)&(X63>-98.
5)&(X105>-89.5)&(X61>-88.0) then 2 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7<-65.
0) then 2 with depth 8

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7>-65.
0)&(X24<-75.0) then 2 with depth 9

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7>-65.
0)&(X24>-75.0)&(X13<-73.0) then 3 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7>-65.
0)&(X24>-75.0)&(X13>-73.0)&(X60<-59.5) then 3 with depth 11

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7>-65.
0)&(X24>-75.0)&(X13>-73.0)&(X60>-59.5)&(X84<-50.0) then 2 with depth 12

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100<-55.5)&(X7>-65.
0)&(X24>-75.0)&(X13>-73.0)&(X60>-59.5)&(X84>-50.0) then 5 with depth 12

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97<-53
.0)&(X173<-37.5) then 3 with depth 9

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97<-53
.0)&(X173>-37.5)&(X50<-57.5) then 5 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97<-53.0)&(X173>-37.5)&(X50>-57.5) then 2 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125<-150.0) then 3 with depth 9

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126<-171.0) then 2 with depth 10

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19<-103.0) then 4 with depth 11

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19>-103.0)&(X33<-132.5) then 2 with depth 12

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19>-103.0)&(X33>-132.5)&(X78<-63.5) then 1 with depth 13

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19>-103.0)&(X33>-132.5)&(X78>-63.5)&(X40<-93.5) then 5 with depth 14

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19>-103.0)&(X33>-132.5)&(X78>-63.5)&(X40>-93.5)&(X5<-74.0) then 4 with depth 15

&(X171>-118.0)&(X163<-98.5)&(X178>-99.5)&(X176>-88.5)&(X101>-68.5)&(X100>-55.5)&(X97>-53.0)&(X125>-150.0)&(X126>-171.0)&(X19>-103.0)&(X33>-132.5)&(X78>-63.5)&(X40>-93.5)&(X5>-74.0) then 1 with depth 15

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172<-262.5) then 1 with depth 5

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174<-151.5) then 1 with depth 6

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130<-118.5) then 4 with depth 7

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6<-67.5) then 2 with depth 8

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6>-67.5)&(X21<-76.5) then 5 with depth 9

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6>-67.5)&(X21>-76.5)&(X131<-86.5) then 4 with depth 10

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6>-67.5)&(X21>-76.5)&(X131>-86.5)&(X137<-47.0) then 4 with depth 11

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6>-67.5)&(X21>-76.5)&(X131>-86.5)&(X137>-47.0)&(X66<-33.0) then 1 with depth 12

&(X171>-118.0)&(X163>-98.5)&(X164<-78.5)&(X172>-262.5)&(X174>-151.5)&(X130>-118.5)&(X6>-67.5)&(X21>-76.5)&(X131>-86.5)&(X137>-47.0)&(X66>-33.0) then 4 with depth 12

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170<-141.0) then 1 with depth 6

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169<-115.5) then 1 with depth 7

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16<-132.0) then 1 with depth 8

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22<-65.5) then 2 with depth 9

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64<-84.0) then 4 with depth 10

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64>-84.0)&(X52<-76.0) then 3 with depth 11

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64>-84.0)&(X52>-76.0)&(X14<-94.5) then 2 with depth 12

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64>-84.0)&(X52>-76.0)&(X14>-94.5)&(X15<-93.5) then 1 with depth 13

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64>-84.0)&(X52>-76.0)&(X14>-94.5)&(X15>-93.5)&(X20<-59.0) then 2 with depth 14

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140<-80.5)&(X170>-141.0)&(X169>-115.5)&(X16>-132.0)&(X22>-65.5)&(X64>-84.0)&(X52>-76.0)&(X14>-94.5)&(X15>-93.5)&(X20>-59.0) then 4 with depth 14

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136<-63.5)&(X168<-106.0) then 1 with depth 7

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83<-51.0)&(X28>-28.0)&(X29>-20.0)&(X38>-21.0) then 5 with depth 14

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59<-39.5)&(X51<-48.0) then 2 with depth 13

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59<-39.5)&(X51>-48.0)&(X27<-26.0) then 2 with depth 14

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59<-39.5)&(X51>-48.0)&(X27>-26.0)&(X30<-29.5) then 3 with depth 15

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59<-39.5)&(X51>-48.0)&(X27>-26.0)&(X30>-29.5) then 5 with depth 15

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158<59.5) then 3 with depth 13

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74<-223.0) then 2 with depth 14

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37<-140.5) then 2 with depth 15

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41<-105.5) then 5 with depth 16

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41>-105.5)&(X35<-119.0) then 2 with depth 17

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41>-105.5)&(X35>-119.0)&(X76<-190.0) then 2 with depth 18

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41>-105.5)&(X35>-119.0)&(X76>-190.0)&(X77<-162.5) then 3 with depth 19

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41>-105.5)&(X35>-119.0)&(X76>-190.0)&(X77>-162.5)&(X68<-92.5) then 3 with depth 20

&(X171>-118.0)&(X163>-98.5)&(X164>-78.5)&(X140>-80.5)&(X136>-63.5)&(X128>-66.0)&(X129>-57.5)&(X150>-23.5)&(X159>13.0)&(X83>-51.0)&(X59>-39.5)&(X158>59.5)&(X74>-223.0)&(X37>-140.5)&(X41>-105.5)&(X35>-119.0)&(X76>-190.0)&(X77>-162.5)&(X68>-92.5) then 1 with depth 20