David D'Attile
Prof. Kauchak
CS158
20 February 2022

Assignment 05: OVA vs. AVA

1. *Train a decision tree classifier on ALL of the data and set the depth limit at 5. Use the toString method to print out the learned tree. Look at the words. Do they make sense? Do any of them stand out? Include the tree along with 2-3 sentences describing what you see:*

```
(chardonnay
 (pinot
  (blanc
   (zinfandel
    (merlot
     predict=1.0
     predict=7.0)
    (sauvignon
     predict=10.0
     predict=15.0))
   (sauvignon
    (a
     predict=16.0
     predict=17.0)
    (young
     predict=0.0
     predict=15.0)))
  (grigio
   (gris
    (brut
     predict=2.0
     predict=6.0)
    predict=3.0)
   (valley
    predict=4.0
    predict=15.0)))
 (pinot
  (bubbles
   (brut
    (sparkling
     predict=5.0
     predict=6.0)
    predict=6.0)
   predict=6.0)
  (syrah
   (any
    (aromatics
     predict=6.0
     predict=5.0)
    predict=4.0)
   predict=2.0)))
```

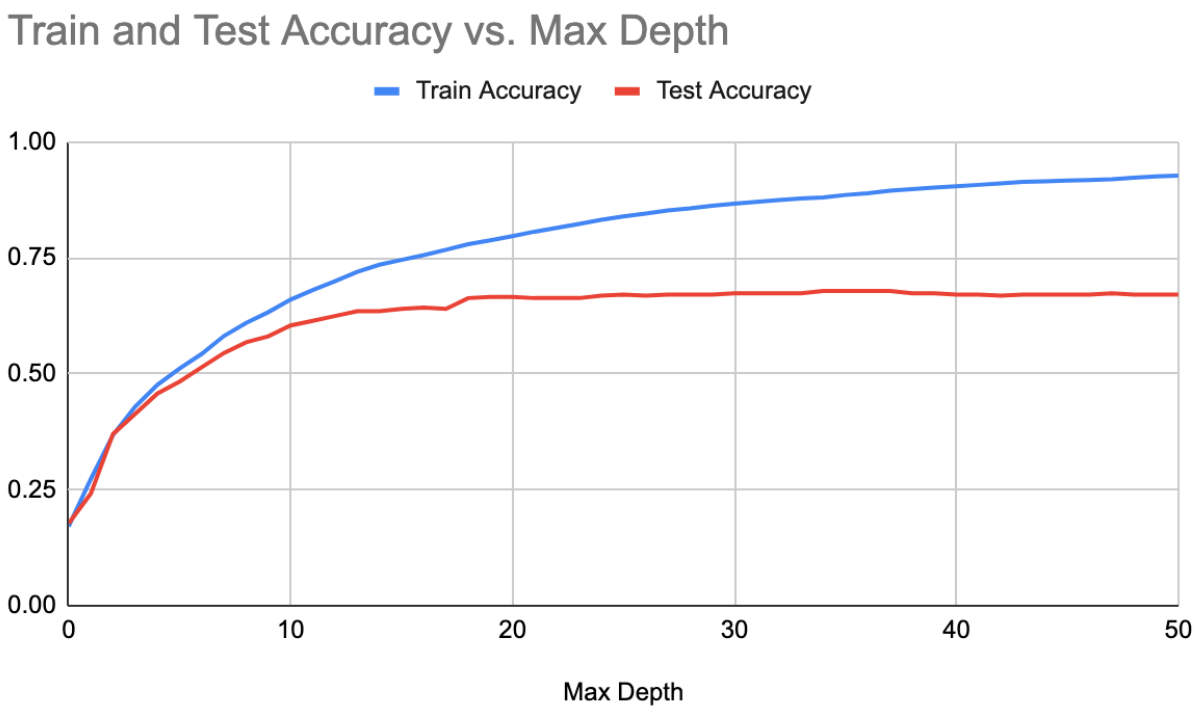
Description:

When we set the depth limit to a low number such as 5 for this data set, we see that many of the most important features for the tree to split on are the actual names of wines such as “pinot”, “chardonnay”, etc. This seems reasonable to me since I would consider seeing the name of a wine in its description as a fairly high indicator

2. *If you just predicted the majority class, what would the accuracy be? This is often a reasonable baseline to measure against (hopefully we can beat this!).*

   The majority class is Cabernet-Sauvignon (label 1.0). If we just predicted this majority label, we would be correct 17.22% of the time (335 / 1,945).

3. *On a random 80/20 split of the data, train the decision tree classifier and evaluate both the training and testing accuracy for depth limits ranging from 0 to 50 (for consistency, all tests should be run on exactly the SAME split). Include the output of your results in either a graph or table. What is the best depth to use? Do you see evidence of overfitting?*



Based on the chart above, we see heavy overfitting at a max depth of approximately 10. Based on this overfitting tendency, I would argue that the best max depth hyperparameter to use for this model is 10 even though higher test accuracy was achieved in the 30 - 40 max depth range.

4. *On a 10-fold cross validation of the wine data set (use the same 10-fold split for all three), calculate accuracies for the following:*
    - *OVA with decision trees sized 1, 2 and 3.*
    - *AVA with decision trees sized 1, 2 and 3.*
    - *Multiclass decision tree (i.e. just by itself) with your best limit found*

    Put these all in a spreadsheet or other format (you'll have 70 numbers, 10 for each experiment). Run a t-test to validate which approach is best. What is the best approach? Is this surprising? Include your table of results and a short write-up describing your results:

Fold	OVA-1	OVA-2	OVA-3	AVA-1	AVA-2	AVA-3	DT-3
0	0.525773	0.56701	0.556701	0.56701	0.592784	0.592784	0.35567
1	0.608247	0.634021	0.649485	0.608247	0.649485	0.654639	0.474227
2	0.56701	0.597938	0.608247	0.57732	0.618557	0.654639	0.443299
3	0.56701	0.613402	0.603093	0.556701	0.670103	0.664948	0.371134
4	0.587629	0.618557	0.603093	0.613402	0.680412	0.680412	0.448454
5	0.603093	0.597938	0.608247	0.608247	0.675258	0.690722	0.443299
6	0.587629	0.639175	0.649485	0.654639	0.670103	0.664948	0.474227
7	0.520619	0.536082	0.551546	0.525773	0.628866	0.639175	0.360825
8	0.556701	0.608247	0.613402	0.572165	0.618557	0.659794	0.407216
9	0.628141	0.648241	0.678392	0.638191	0.683417	0.703518	0.452261
Average	0.5751852	0.6060611	0.6121691	0.5921695	0.6487542	0.6605579	0.4230612
T-Tests:							
AVA-3 vs. OVA-1	0.0000003014821062						
AVA-3 vs. OVA-2	0.00005369061483						
AVA-3 vs. OVA-3	0.0004648709531						
AVA-3 vs. AVA-1	0.0001108117182			`			
AVA-3 vs. AVA-2	0.04954470329						
AVA-3 vs. DT-3	0.000000005916549328						

Based on the results above, we can see that the AVA w/DT (max depth 3) is the best approach for the wine dataset. I do not find this surprising given that the AVA algorithm makes predictions based on the weighted vote of confidence among all

classifiers c(i,j) where “c” is a classifier, “i” and “j” are labels, and each example to be classifier is compared against every (i,j) pairing of labels in the dataset. From the T-Test results, we can see that there is not much confidence in the effectiveness of AVA-3 vs. AVA-2, but contrastingly, there is a significant performance improvement when comparing AVA-3 to any OVA algorithm or the basic decision tree algorithm with a maximum depth of 3.

5. *Time both OVA and AVA on some reasonable test of the wine data set and measure both training time and testing time (separately). You can use the ClassifierTimer class if you'd like or just do it yourself. Do the timings make sense? Include your results and 2-3 sentences describing how you generated the timings and explaining the results:*

Classifier	Train Time	Classify Time
OVA w/DT-3	28.505	0.0013
AVA w/DT-3	20.7777	0.008
OVA w/Avg. P-10	2.1483	0.02233
AVA w/Avg. P-10	3.277	0.2973

The timing above was implemented with ClassifierTimer and 3 runs per classifier that were eventually averaged together. We can see that when compared to OVA, AVA takes apprx. ⅔ as much time to train, but nearly 7 times longer to classify. I believe these results make sense since to train AVA, we break the training data into binary subsets of itself compared to OVA where we simply binarize the whole data set at each iteration.

As an aside, I timed the OVA and AVA algorithms with the Avg. Perceptron ClassifierFactory, and using these classifiers led to *much* faster training/testing speeds. I was not able to compare these implementation's accuracies to the DT implementation, but the speed difference alone is worth recognizing.

6. *On this data set, what would you say is the best approach to use? Briefly justify your answer:*

Given the performance and timing results, I believe that AVA is the best classifier algorithm for the provided wines dataset. It is faster to train, and even though it takes longer to classify, it is measurably better when compared to the OVA algorithm for all DT hyper parameter values for this data set which we saw in Q4.

7. *Train the OVA classifier with decision trees sized 3 on all of the data. What is the tree for the zinfandel classifier (again, use the toString method to print it out)? Does this make sense? What words are indicative of the class? What words are not indicative of the class? Include a short summary describing your analysis of the tree:*

Tree:

```
(zinfandel
        (zin
                (amador
                        predict=-1.0
                        predict=1.0)
                predict=1.0)
        (sauvignon
                (grenache
                        predict=1.0
                        predict=-1.0)
                predict=-1.0))
```

Description:

Here, we see that the words “zinfandel”, “zin”, and “amador” are the most indicative of the class. Contrastingly, the words “sauvignon”, “grenache”, and the *lack* of “zinfandel” are not indicative. These results make sense since seeing “zinfandel” or “zin” would be a hefty indicator that we are talking about zinfandel wine. Similarly, seeing the word “sauvignon” sways the prediction away from positive since it is an entirely different kind of wine, and seeing the word “grenache” also sways the prediction away since zinfandel wine is not made from grapes from the grenache plant.