CMSC422 Project 2 Writeup

Arjun Baradwaj

Yoshi Fujimoto

Prayaag Venkat

WU1:

OAA

(A) After training a one against all multiclassifier on the small data set using depth 3 DTs:

```
h = multiclass.OAA(5, lambda: DecisionTreeClassifier(max_depth=3))

h.train(WineDataSmall.X, WineDataSmall.Y)

util.showTree(h.f[0], WineDataSmall.words)
```

We obtain the following description of the DT that classifies examples as label 0 (Sauvignon-Blanc) or not label 0, corresponding to classes 1 and 0 respectively:

```
citrus?
-N-> lime?
|    -N-> gooseberry?
|    |    -N-> class 0  (356 for class 0, 10 for class 1)
|    |    -Y-> class 1  (0 for class 0, 4 for class 1)
|    -Y-> hint?
|    |    -N-> class 1  (1 for class 0, 15 for class 1)
|    |    -Y-> class 0  (2 for class 0, 0 for class 1)
-Y-> grapefruit?
|    -N-> flavors?
|    |    -N-> class 1  (4 for class 0, 12 for class 1)
|    |    -Y-> class 0  (11 for class 0, 5 for class 1)
|    -Y-> lingering?
|    |    -N-> class 1  (0 for class 0, 14 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
```

First, observe that in examples where the word “citrus” was present, (31/47) * 100 ~ 66% were classified as Sauvignon-Blanc, and in examples where “citrus” is absent, (29/388) * 100 ~ 7.5% were classified as Sauvignon-Blanc. Since this is the first (and thus best differentiating) split in the DT, this indicates that the word citrus is most indicative of Sauvignon-Blanc. Next, for examples where the word “citrus” was absent, but the word “lime” was present, a similar computation to the one above shows that a significantly larger fraction of the examples with “lime” were classified Sauvignon-Blanc, while a much lower fraction of examples in which “lime” was absent were Sauvignon-Blanc. Again, repeating this analysis for the first few levels of the above DT shows that the words “citrus”, “lime”, and “grapefruit” were most indicative of Sauvignon-Blanc. Words that are indicative of not being Sauvignon-Blanc are “hint”, “flavors”, and “lingering”, as seen in the lower levels of the above tree. Presence of these words is classified as not Sauvignon-Blanc. Additionally, the words found at the higher levels of DTs of other classes (that are not Sauvignon-Blanc) will also be indicative of not being Sauvignon-Blanc because they separate a single class from other classes, one of which is Sauvignon-Blanc.

For Pinot-Noir, we can generate a similar depiction of the DT:

```
cherry?
-N-> raspberries?
|     -N-> strawberry?
|     |     -N-> class 0  (225 for class 0, 58 for class 1)
|     |     -Y-> class 1  (0 for class 0, 4 for class 1)
|     -Y-> lingering?
|     |     -N-> class 1  (0 for class 0, 12 for class 1)
|     |     -Y-> class 0  (1 for class 0, 0 for class 1)
-Y-> cassis?
|     -N-> petit?
|     |     -N-> class 1  (36 for class 0, 68 for class 1)
|     |     -Y-> class 0  (8 for class 0, 0 for class 1)
|     -Y-> allspice?
|     |     -N-> class 0  (21 for class 0, 0 for class 1)
|     |     -Y-> class 1  (0 for class 0, 2 for class 1)
```

Again, using an analysis identical to the one above, we find that words indicative of Pinot-Noir are “cherry”, “raspberries”, and “strawberry”, while words indicative of not being Pinot Noir are “cassis”, “petit”, and “lingering”.

(B)

Time taken to train depth tree decision trees on the full WineData set is 2.74699997902 seconds, and the test accuracy obtained is 0.36827458256.

```
peaches?
-N-> nectarine?
|     -N-> chilled?
|     |     -N-> class 0  (1036 for class 0, 1 for class 1)
|     |     -Y-> class 0  (6 for class 0, 1 for class 1)
|     -Y-> edge?
|     |     -N-> class 0  (13 for class 0, 1 for class 1)
|     |     -Y-> class 1  (0 for class 0, 1 for class 1)
-Y-> milk?
|     -N-> d?
|     |     -N-> class 0  (14 for class 0, 0 for class 1)
|     |     -Y-> class 1  (0 for class 0, 1 for class 1)
|     -Y-> class 1        (0 for class 0, 3 for class 1)
```

Words indicative of Viognier are “peaches”, “milk”, using the same reasoning described above.

(C)

Using same dataset and models from (B) above, accuracy with confidence is 0.37940630797773656, whereas accuracy with zero-one predictions is 0.25139146567717996. So, using zero-one predictions reduces test accuracy.

AVA:

(A)

To find the words most indicative of Sauvignon-Blanc, we considered the DTs that were trained to classify Sauvignon-Blanc vs. another class. If we analyze each of these trees as we did above, and then we take the most frequently appearing words across all the DTs, it follows that these words are most indicative of Sauvignon-Blanc. As partially shown in a small sample of the trees below, the words most indicative of Sauvignon-Blanc are “citrus”, “lime”, and “crisp”. Words most indicative of not being Sauvignon-Blanc are “dark”, “tannins”, “mild”, and “fruity”. Words indicative of Pinot-Noir are “cherry”, “cherries”, and “red”. Words indicative of not being Pinot-Noir are “apple”, “purple”, “grape”, “peach”, and “peaches”.

```
>>> util.showTree(h.f[1][0], WineDataSmall.words)
citrus?
-N-> lime?
|     -N-> refreshing?
|     |     -N-> class 0   (187 for class 0, 9 for class 1)
|     |     -Y-> class 1   (0 for class 0, 5 for class 1)
|     -Y-> class 1         (0 for class 0, 15 for class 1)
-Y-> class 1     (0 for class 0, 31 for class 1)
>>> util.showTree(h.f[2][0], WineDataSmall.words)
crisp?
-N-> lime?
|     -N-> lemon?
|     |     -N-> class 0   (141 for class 0, 9 for class 1)
|     |     -Y-> class 1   (0 for class 0, 8 for class 1)
|     -Y-> meats?
|     |     -N-> class 1   (0 for class 0, 13 for class 1)
|     |     -Y-> class 0   (1 for class 0, 0 for class 1)
-Y-> red?
|     -N-> class 1         (0 for class 0, 30 for class 1)
|     -Y-> class 0         (2 for class 0, 0 for class 1)
```

```
>>> util.showTree(h.f[3][0], WineDataSmall.words)
thai?
-N-> very?
|     -N-> another?
|     |     -N-> class 1  (4 for class 0, 56 for class 1)
|     |     -Y-> class 0  (1 for class 0, 0 for class 1)
|     -Y-> ripe?
|     |     -N-> class 1  (1 for class 0, 4 for class 1)
|     |     -Y-> class 0  (4 for class 0, 0 for class 1)
-Y-> class 0    (5 for class 0, 0 for class 1)
>>> util.showTree(h.f[4][0], WineDataSmall.words)
apple?
-N-> pasta?
|     -N-> quite?
|     |     -N-> class 1  (11 for class 0, 56 for class 1)
|     |     -Y-> class 0  (3 for class 0, 0 for class 1)
|     -Y-> class 0       (4 for class 0, 0 for class 1)
-Y-> bright?
|     -N-> class 0       (10 for class 0, 0 for class 1)
|     -Y-> reminiscent?
|     |     -N-> class 1  (0 for class 0, 4 for class 1)
|     |     -Y-> class 0  (1 for class 0, 0 for class 1)
```

(B)

The time taken was 7.10199999809 seconds and the accuracy obtained was 0.270871985158. Words indicative of Viognier are “peaches”, “peach”, “milk”, “nectarine”. These words were the most frequent words that appeared at the top of each of the DTs that classified Viognier vs another class.

(C)

Using zero-one for predictions, the accuracy is 0.26437847866419295. When using confidence for predictions, the accuracy was 0.2717996289424861. So, in this case, it does not make a significant difference when using zero-one vs. confidence for predictions.

WU2:

The test accuracy for a balanced tree on WineData when using a DecisionTreeClassifier with max depth 3 is 0.31076066790352502, as shown below.

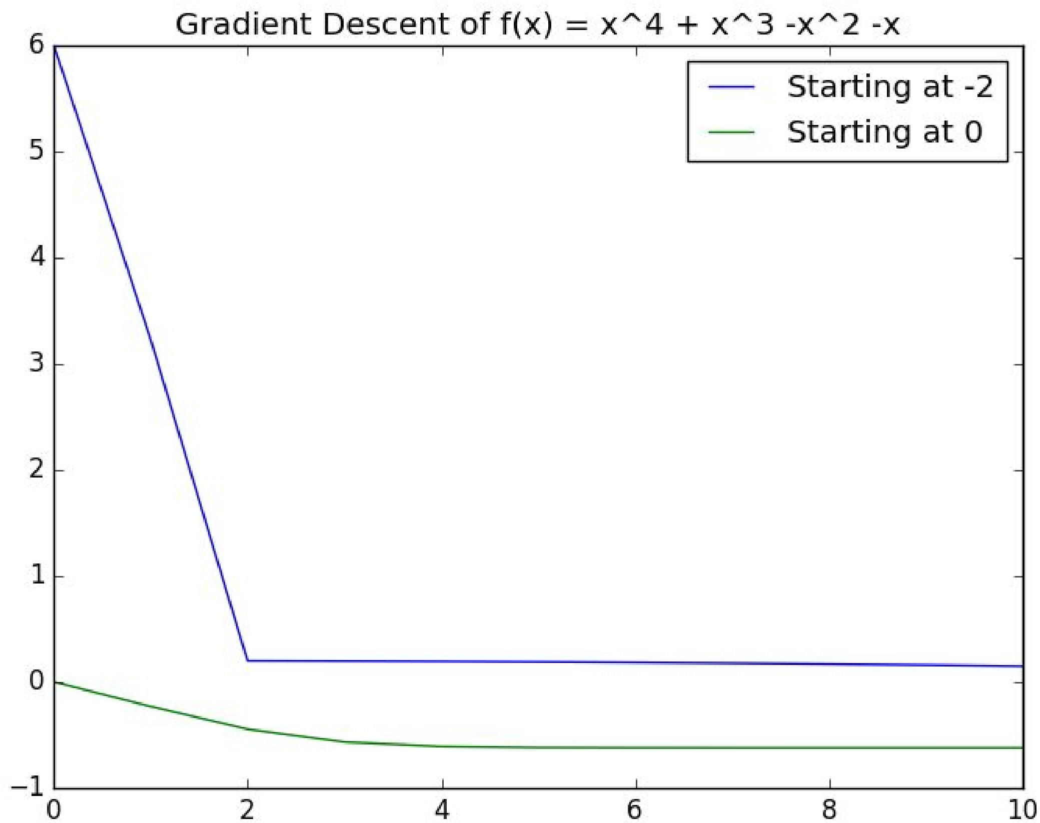
```
>>> t = multiclass.makeBalancedTree(range(20))
>>> h = multiclass.MCTree(t, lambda: DecisionTreeClassifier(max_depth=3))
>>> h.train(WineData.X, WineData.Y)

training classifier for [0, 1, 2, 3, 4, 5, 6, 7, 8, 9] versus [10, 11, 12, 13, 14, 15, 16, 17, 18, 19]
training classifier for [0, 1, 2, 3, 4] versus [5, 6, 7, 8, 9]
training classifier for [0, 1] versus [2, 3, 4]
training classifier for [0] versus [1]
training classifier for [2] versus [3, 4]
training classifier for [3] versus [4]
training classifier for [5, 6] versus [7, 8, 9]
training classifier for [5] versus [6]
training classifier for [7] versus [8, 9]
training classifier for [8] versus [9]
training classifier for [10, 11, 12, 13, 14] versus [15, 16, 17, 18, 19]
training classifier for [10, 11] versus [12, 13, 14]
training classifier for [10] versus [11]
training classifier for [12] versus [13, 14]
training classifier for [13] versus [14]
training classifier for [15, 16] versus [17, 18, 19]
training classifier for [15] versus [16]
training classifier for [17] versus [18, 19]
training classifier for [18] versus [19]
>>> P = h.predictAll(WineData.Xte)
>>> mean(P == WineData.Yte)
0.31076066790352502
```

WU3:

With the function, f = x^2, and 100 iterations negative step sizes such as -2 or -1 diverge. A step size of 0 is constant while 0.5 and 1 converge rather quickly. Larger positive step sizes converge much later or diverge. As you get to a step size of 7, it no longer converges within 100 iterations.

WU4:

The function f(x) = x^4 + x^3 - x^2 - x has a local minimum at x = -1 and a global minimum at x = 0.6404. When starting at x = -2, the function gets caught at the local minimum, but starting at x = 0, the function converges at the global minimum.



WU5:

The logistic loss function worked the best, with both the highest training and test accuracy. This means the model did not overfit the training data, yet was still able to generalize properly and hit approximately 97.4+% test accuracy.

```
Logistic:
Training accuracy 0.995951, test accuracy 0.97417
Printing WineDataBinary.words:
['yellow', 'four', 'elegant', 'contributes', 'herb', 'jasmine',
Words array length: 819

Printing small items:
tannins
dark
blackberry
black
cherry


Printing large items:
crisp
tropical
acidity
lime
citrus
```

Result from console, from a script written to sort, and detect values for map (in reverse order)

Top 5 Positive:

1. Citrus (Most positive, at 0.883289753118)
2. Lime
3. Acidity
4. Tropical
5. Crisp (5th most positive, at 0.606423261902)

Top 5 Negative:

1. Tannins (Most negative, at -1.1695212164)
2. Dark
3. Blackberry
4. Black
5. Cherry (5th most negative, at -0.532191672468)

The words in the positive set and negative set are nearly perfectly polar, which makes sense given the high training and test accuracy. The top 5 positive words describe acidity and sour fruits, as well as a tropical, crisp, wine texture. On the other hand, in the negatives are “dark”, “black”, fruits including “blackberry” and “cherry”, and “tannins”, which is a property that describes the dry texture of a wine. The words within a corresponding group are closely related to each other, yet there is a clear divide between the two groups on opposite ends of the predicted weight spectrum.