Project2 Write Up

#WU1
For OAA:
(A). We trained OAA multiclass classifier on the small data set using depth 3 DTs and obtained the following result.

```
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-> apple?
|    |    -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)
```

We observed that in examples where the word "citrus" was present, (31/47)*100% ~66% were classified as Sauvignon-Blanc. But in the examples where "citrus" was absent, (29/388)*100~7.5% were classified as Sauvignon-Blanc. In addition to this, since "citrus" was the split on the top of the DT which means the "citrus" is the best differentiating feature, so that we conclude the word citrus is the most indicative of Sauvignon-Blanc. If we consider the case where the word "citrus" was absent and the word "lime" was present, there were (15/18)*100~83% examples classified as Sauvignon-Blanc. And there were (14/270)*100~5.2% examples classified as Sauvignon-Blanc if "lime" was absent. We found a similar pattern with word "grapefruit". Repeating this analysis for the first few levels of the above DT shows that the words "citrus", "limes", and "grapefruit" were most indicative of being Sauvignon-Blanc. We can verify this as those words appear at the lower levels of the tree (near the top of the tree).

Then, we observed that when "apple" was present, 100% examples were classified as not Sauvignon-Blanc. But if "apple" was absent, there were 6.25% examples were classified as not Sauvignon-Blanc. When we apply the similar calculation to "flavors" and "lingering", we saw that both of them gave a larger fraction of examples of not being Sauvignon-Blanc with the word was present, and a smaller fraction of examples of being not Sauvignon-Blanc with the word was absent. Thus, "apple", "flavors" and "lingering" are most indicative of not being Sauvignon-Blanc.

We ran the command to get the tree for Pinot-Noir:

```
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-> cocoa?
|    |    -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)
```

We did the analysis identical to the one above and found that the words are indicative of being Pinot-Noir were "cherry", "raspberries", and "strawberry" which gave the percentage of 51.9%, 92% and 100% of being Pinot-Noir with the present of those words. The words "cassis", "petit" and "cocoa" gave a larger fraction of not being Pinot-Noir, so those words are most indicative of not being Pinot-Noir.

(B). The test accuracy is 0.36734693877551022 and it takes 0.586641788482666 seconds to train.

The tree we get is the following.

```
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-> savory?
|    |    -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-> harmonious?
|    |    -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 "peaches" and "milk" are most indicative of this since they show a larger fraction of being Viognire when they are present compared to when they are absent.

(C). Using the same dataset and models from (B), accuracy with confidence is 0.36827458256029683 and with zero-one prediction is 0.24304267161410018. The difference is 0.36827458256029683 - 0.24304267161410018 =

0.12523191094619665. We concluded that using confidence is 0.12523191094619665 better than using the zero-one prediction in accuracy.

For AVA:

(A). To consider 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 gives the words are most indicative of Sauvignon-Blanc. We trained the AVA multiclass classifier on small sample with depth 3 and obtained the following result.

```
>>> 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-> ginger?
|    |    -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-> warm?
|    |    -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-> white?
|    |    -N-> class 1  (0 for class 0, 4 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
```

Therefore, we conclude that the words are most indicative of being Sauvignon-Blanc are “citrus”, “lime” and “crisp” since the it gave a larger fraction of being Sauvignon-Blanc when those words were present. The words are most indicative of not being Sauvignon-Blanc are “meats”, “red” and “ripe” as it gave a larger fraction of not being Sauvignon-Blanc.

```
>>> 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-> roasted?
|    |    -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[2][1], WineDataSmall.words)
cassis?
-N-> acidity?
|    -N-> salmon?
|    |    -N-> class 1  (92 for class 0, 129 for class 1)
|    |    -Y-> class 0  (11 for class 0, 0 for class 1)
|    -Y-> tannins?
|    |    -N-> class 0  (22 for class 0, 0 for class 1)
|    |    -Y-> class 0  (15 for class 0, 11 for class 1)
-Y-> tea?
|    -N-> 100?
|    |    -N-> class 1  (1 for class 0, 47 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
|    -Y-> class 0       (2 for class 0, 0 for class 1)
```

```
>>> util.showTree(h.f[3][2], WineDataSmall.words)
crisp?
-N-> peach?
|    -N-> pear?
|    |    -N-> class 1  (3 for class 0, 142 for class 1)
|    |    -Y-> class 0  (2 for class 0, 0 for class 1)
|    -Y-> class 0       (3 for class 0, 0 for class 1)
-Y-> red?
|    -N-> class 0       (7 for class 0, 0 for class 1)
|    -Y-> class 1       (0 for class 0, 2 for class 1)
>>> util.showTree(h.f[4][2], WineDataSmall.words)
straw?
-N-> crisp?
|    -N-> shellfish?
|    |    -N-> class 1  (8 for class 0, 142 for class 1)
|    |    -Y-> class 0  (2 for class 0, 0 for class 1)
|    -Y-> red?
|    |    -N-> class 0  (7 for class 0, 0 for class 1)
|    |    -Y-> class 1  (0 for class 0, 2 for class 1)
-Y-> class 0    (12 for class 0, 0 for class 1)
```

With the same reasoning, the words are most indicative of being Pinot-Noir are “red” and “roasted”. The words are not indicative of not being Pinot-Noir are “lemon”, “cassis”, and “peach”.

(B). The test accuracy we obtained is 0.26530612244897961. And the time it took to train is 0.6480460166931152 seconds.

Run the AVA multiclass classifier with full dataset and depth 3, we get the following (partial) result.

```
>>> util.showTree(h.f[17][0], WineData.words)
floral?
-N-> enjoy?
|    -N-> thai?
|    |    -N-> class 1  (2 for class 0, 59 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
|    -Y-> class 0       (1 for class 0, 0 for class 1)
-Y-> delicate?
|    -N-> class 0       (4 for class 0, 0 for class 1)
|    -Y-> class 1       (0 for class 0, 1 for class 1)
>>> util.showTree(h.f[17][1], WineData.words)
peaches?
-N-> peach?
|    -N-> pear?
|    |    -N-> class 1  (0 for class 0, 187 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 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)
>>> util.showTree(h.f[17][2], WineData.words)
peach?
-N-> peaches?
|    -N-> pear?
|    |    -N-> class 1  (0 for class 0, 144 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 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)
>>> util.showTree(h.f[17][3], WineData.words)
peaches?
-N-> nectarine?
|    -N-> tannin?
|    |    -N-> class 1  (1 for class 0, 15 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
|    -Y-> class 0       (2 for class 0, 0 for class 1)
-Y-> class 0    (4 for class 0, 0 for class 1)
>>>
```

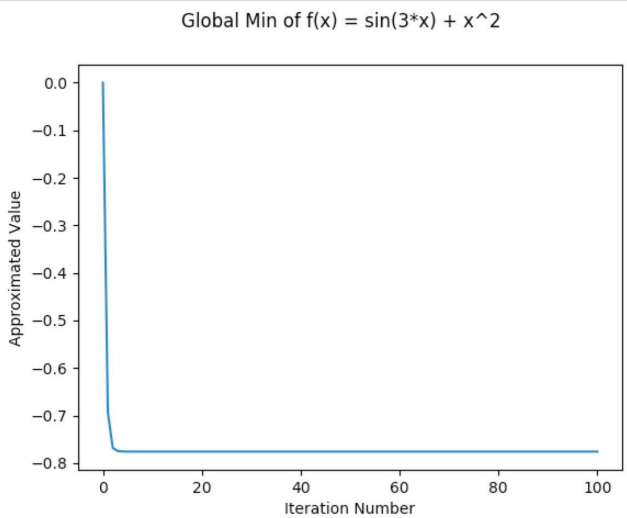
The words that are indicative of Viognier are “floral”, “peaches” and “nectarine” using the same analysis as before.

(C). The test accuracy using confidence is 0.26345083487940629 and using the zero-one prediction is 0.26345083487940629, which is about the same as using confidence. Based on calculation, the difference is 0.

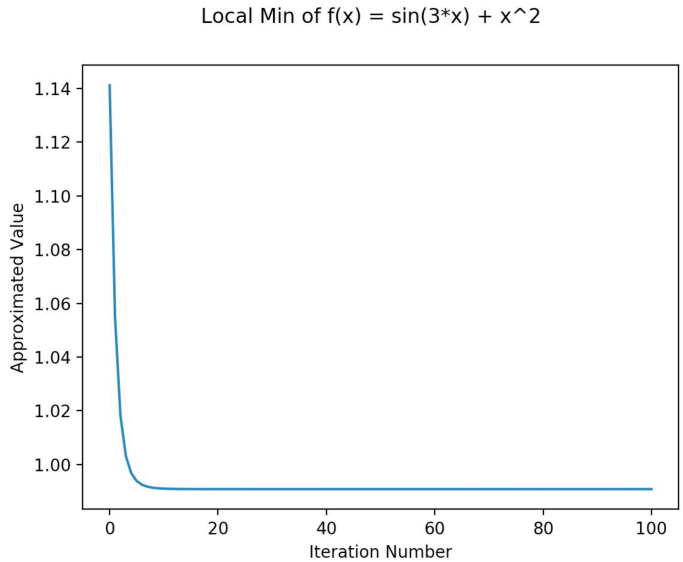
#WU2
The test accuracy we got with a balanced tree on the WineData using a DecisionTreeClassifier with max depth 3 is 0.30890538033395176~30.9%

#WU3
For negative values of step size like -1, the gradient decent algorithm diverges on the function f(x) = x^2. For the step size of 0, the algorithm remains constant at that value. For positive step sizes like 0.5 and 1, algorithm converges rather quickly. Larger positive step sizes converge much later.

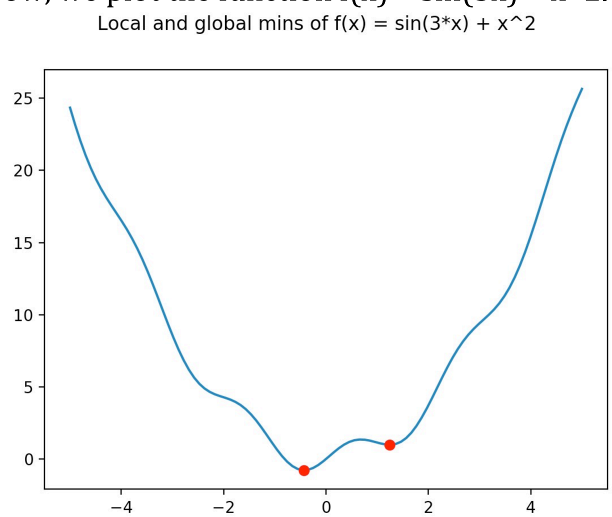
#WU4
One example of non-convex univariate function is f(x) = sin(3x) + x^2. Running the gradient descent algorithm with 100 iterations and step size 0.1 finds the global minimum at x = -0.4273 if the starting value is 0 (x, trajectory = gd.gd(lambda x: sin(3*x) + x**2, lambda x: 3*cos(3*x) + 2*x, 0, 100, 0.1)).



If we try the starting value 1 (x, trajectory = gd.gd(lambda x: sin(3*x) + x**2, lambda x: 3*cos(3*x) + 2*x, 1, 100, 0.1)), it gets caught in a local minimum at x = 1.245.



Now, we plot the function f(x) = sin(3x) + x^2.



#WU5
We ran the model with lambda = 1, iteration = 100 and step size = 0.5. With SquaredLoss, the training accuracy is 0.242915~24.3%, and test accuracy is 0.313653~31.4%. With LogisticLoss, the training accuracy is 0.995951~99.6% and the test accuracy is 0.97417~97.4%. Using HingeLoss, the training accuracy is 0.753036~75.3%, and the test accuracy is 0.686347~68.6%. As we can see, the LogisticLoss yields the best result. Look at the learned weights,

Top 5 positive weights are [0.60642326190168772, 0.68919900790299682, 0.7108905521536929, 0.77012476915577655, 0.88328975311765578]

The words corresponding to the weights are ['tropical', 'acidity', 'lime', 'crisp', 'citrus']

The positive words are those are most likely to predict white wine, which is positive class.

Top 5 negative weights are [-1.1695212164040423, -0.76530939064270642, -0.68359316778937784, -0.62959072814343398, -0.53219167246753041]

The words corresponding to the weights are ['tannins', 'black', 'dark', 'cherry', 'blackberry']

The negative words are those which are most likely to predict the red wine, which is negative class.