# Multiclass Classification

## WU1

Answer questions A, B and C for both OAA and AVA. (Questions about “indicative” are open-ended. Any reasonable answers with analysis will be credited.)
(A) Train depth 3 decision trees on the WineDataSmall task. What words are most indicative of being Sauvignon-Blanc? Which words are most indicative of not being Sauvignon-Blanc? What about Pinot-Noir (label==2)?
**OAA**: After training a OAA classifier on the WineDataSmall task using depth 3 decision trees (max_depth = 3, k = 5), I obtain the following decision tree that classifies examples as label 0 (likely to be Sauvignon-Blanc) or not (likely not be), as classes 1 and 0 accordingly:

```
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)
```

In examples where the word “citrus” was present, 66% (31/47) were classified as Sauvignon-Blanc, while 7.5% (29/388) of examples are classified as Sauvignon-Blanc when the word was absent, making “citrus” one of the words most indicative for Sauvignon-Blanc.

- For the examples where “citrus” is absent, 83% (15/18) of examples with the word “lime” are classified as 1 while 3.8% (14/370) of examples without the word are classified as 1.
- For the examples where “citrus” is present, 93% (14/15) of examples with the word “grapefruit” are classified as 1 while 53% (17/32) of examples without the word are classified as 1.

According to the computations above and the decision tree, the words most indicative for Sauvignon-Blanc are “grapefruit”, “lime”, and “citrus”, while the words most indicative of an example not being Sauvignon-Blanc are “flavors”, “lingering”, “hint”. Additionally, an example not having the aforementioned most indicative words is also indicative of not being Sauvignon-Blanc.
The following is the decision tree that classifies examples as likely to be Pinot-Noir or not can:
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-> verdot?
|    |    -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)
Using the same computational logic for Sauvignon-Blanc, the words most indicative of an example being Pinot-Noir if present are “cherry”, “raspberries”, and “strawberry”, while the words most indicative of not being if present are “cassis”, “verdot”, and “cocoa”. Similarly, the absence of the most indicative words are also indicative of an example not likely being Pinot-Noir.

**AVA**: After training a AVA classifier on the WineDataSmall task using depth 3 decision trees to classify Sauvignon-Blanc against Cabernet-Sauvignon, Pinot-Noir, Pinot-Gris, and Pinot-Grigio, I obtain the following decision trees that classify examples as label 0 (likely to be Sauvignon-Blanc) or not (likely not be), as classes 1 and 0 accordingly:

Cabernet-Sauvignon vs Sauvignon-Blanc
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)

Pinot-Noir vs Sauvignon-Blanc

```
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-> harmonious?
|    |    -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)
```

Pinot-Gris vs Sauvignon-Blanc

```
thai?
-N-> very?
|    -N-> vines?
|    |    -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)
```

Pinot-Grigio vs Sauvignon-Blanc

```
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-> reminiscent?
|    |    -N-> class 1  (0 for class 0, 4 for class 1)
|    |    -Y-> class 0  (1 for class 0, 0 for class 1)
```

According to the decision trees above and using the same computational logic for the OAA classifier, when training Sauvignon-Blanc against Cabernet-Sauvignon, the words most indicative of an example being Sauvignon-Blanc if present are “citrus”, “lime”, “freshing”;

against Pinot-Noir the most indicative words are “crisp”, “lime”, “lemon”; against Pinot-Gris the indicative words for Sauvignon-Blanc if absent are “thai”, “ripe”, “vines”; and against Pinot-Grigio the indicative words for Sauvignon-Blanc if absent are “apple”, “bright”, “pasta”. In summary, the words most indicative for Sauvignon-Blanc are “citrus”, “lime”, and “crisp”, while the words most indicative of an example not being Sauvignon-Blanc are “thai”, “apple”, “red”.

The following is the decision tree that classifies examples as likely to be Pinot-Noir or not can against Sauvignon-Blanc, Cabernet-Sauvignon, Pinot-Gris, and Pinot-Grigio:

```
Sauvignon-Blanc vs Pinot-Noir
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-> cheeses?
|   |   -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)

Cabernet-Sauvignon vs Pinot-Noir
cassis?
-N-> acidity?
|   -N-> duck?
|   |   -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  (14 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)

Pinot-Gris vs Pinot-Noir
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)
```

```
Pinot-Grigio vs Pinot-Noir
straw?
-N-> crisp?
|   -N-> example?
|   |   -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)
```

Using the same computational logic earlier, when training Pinot-Noir against Sauvignon-Blanc, the words most indicative of an example being Pinot-Noir if present are “crisp”, “lime”, “lemon”; against Cabernet-Sauvignon the most indicative words are “cassis” if “tea” and “100” are absent, or if “cassis”, “acidity”, and “duck” are all absent; against Pinot-Gris the indicative words for Pinot-Noir are “crisp”, or if “crisp”, “peach”, and “pear” are all absent; and against Pinot-Grigio the indicative words for Pinot-Noir is “crisp” if “straw” and “example” are absent. In summary, the words most indicative for Pinot-Noir are “crisp” and “cassis”, while the words most indicative of an example not being Pinot-Noir are “tea”, “acidity”, “duck”, “peach”, “pear”, “straw”, “example”.

(B) Train depth 3 decision trees on the full WineData task (with 20 labels). What accuracy do you get? How long does this take (in seconds)? One of my least favorite wines is Viognier -- what words are indicative of this?

**OAA**: The test accuracy I obtained is 0.3719851576994434, and it took 0.1483900547027588 seconds. According to the decision tree for classifying Viognier I obtained, the words indicative of it are “peaches” and “milk”, using the same computational logic as above.

**AVA**: The test accuracy I obtained is 0.2699443413729128, and it took 0.19247674942016602 seconds. According to the decision trees for classifying Viognier against the other wines I obtained, the words indicative of it are “peaches”, “peach”, “milk”, “floral”.

(C) Compare the accuracy using zero-one predictions versus using confidence. How much difference does it make?

**OAA**: The test accuracy I obtained using zero-one predictions is 0.24304267161410018, about 13% worse than using confidence.
**AVA**: The test accuracy I obtained using zero-one predictions is 0.2588126159554731, about the same as using confidence.

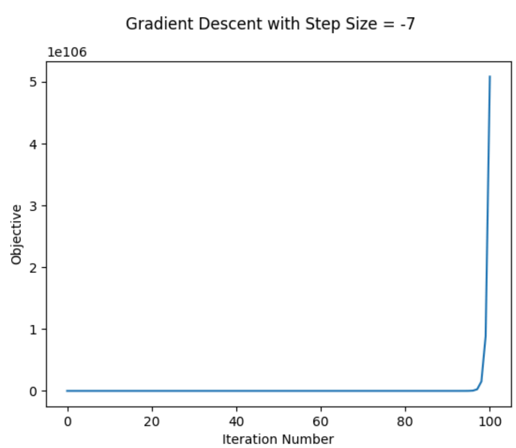
## WU2

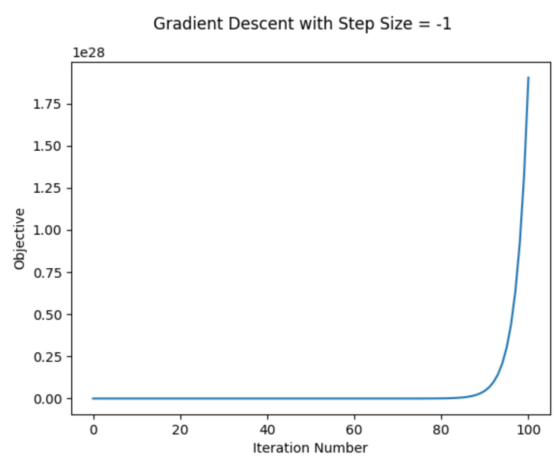
Now, you must implement a tree-based reduction. Show the test accuracy you get with a balanced tree on the WineData using a DecisionTreeClassifier with max depth 3.
The test accuracy with a balanced tree on the WineData dataset using a DecisionTreeClassifier with max depth 3 is around 20%.

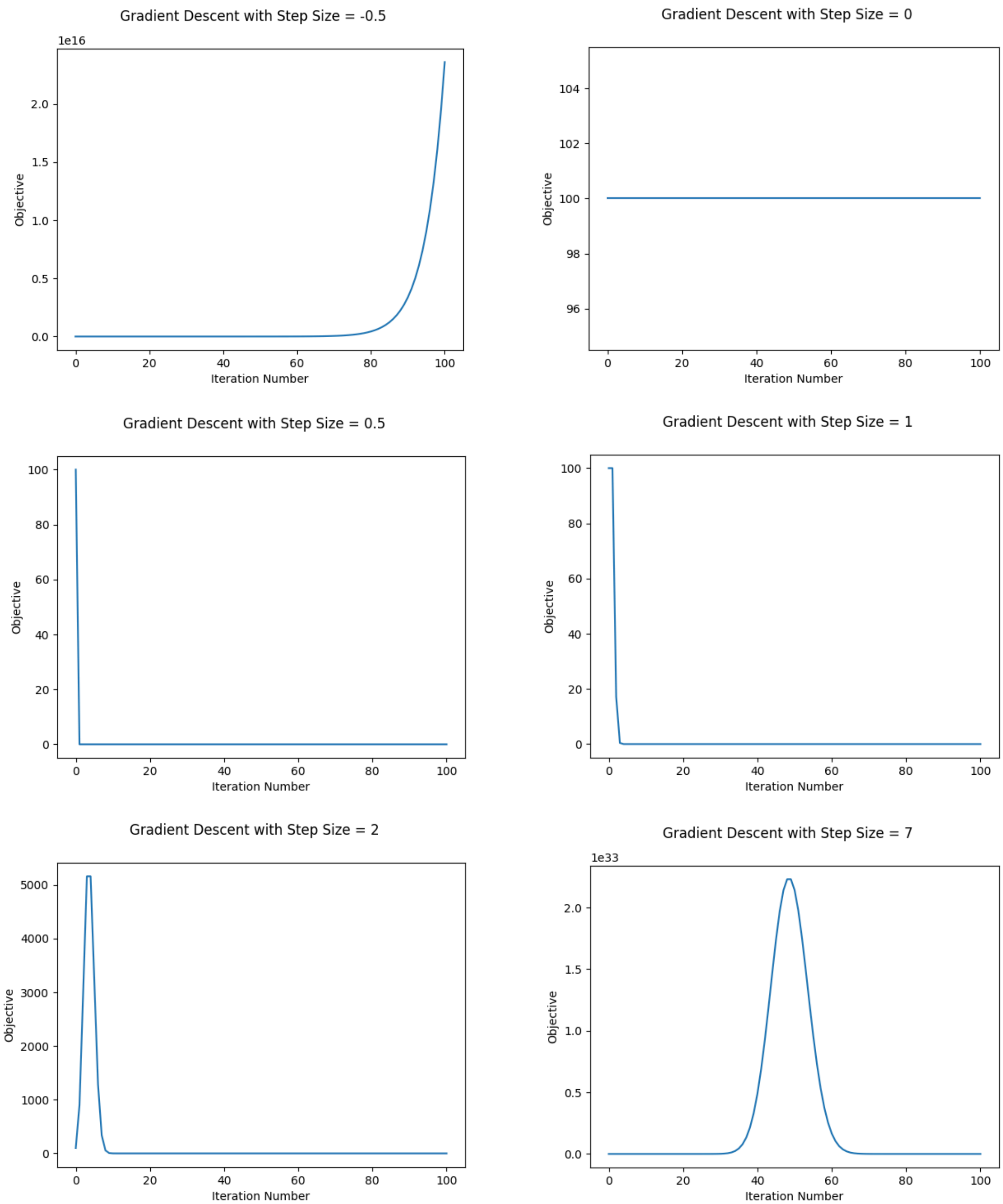
# Gradient Descent

## WU3

What is the impact of the step size on convergence? Find values of the step size where the algorithm diverges and converges.
With the function f(x) = x^2, if step size is negative, the algorithm diverges as iterations increase, with smaller negative numbers diverging less. If step size is 0, the algorithm is constant. If step size is positive but small, the algorithm converges in a few iterations, while if the step size is larger than 1, it briefly diverges after a few iterations before converging. The larger the step size, the later it begins converging after diverging.



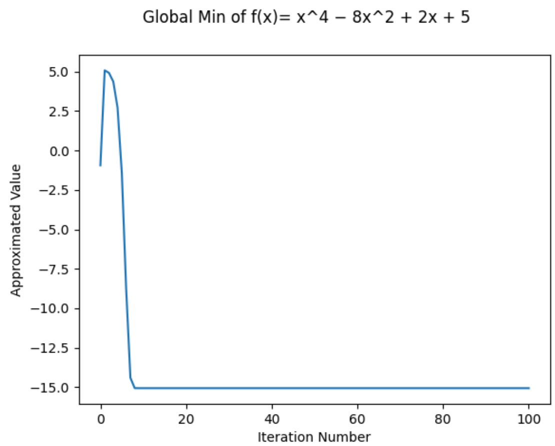



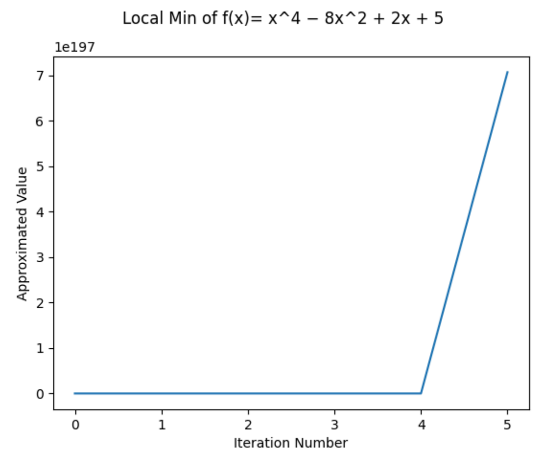


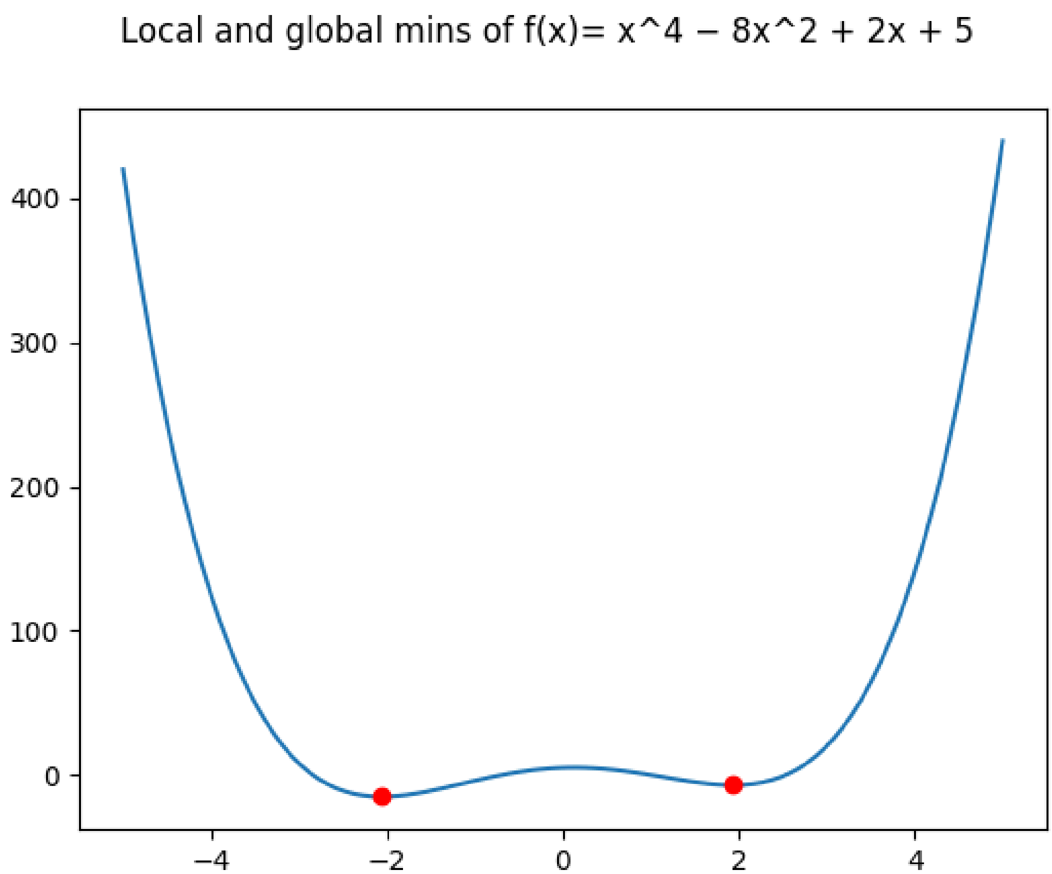

# WU4

Come up with a non-convex univariate optimization problem. Plot the function you're trying to minimize and show two runs of gd, one where it gets caught in a local minimum and one where it manages to make it to a global minimum. (Use different starting points to accomplish this.)

My non-convex univariate optimization problem is f(x)=x^4−8x^2+2x+5. For the gradient descent algorithm, I will use 100 iterations and a step size of 0.1. For the gd that gets caught in a local minima I used 3.5 as the starting point, for the gd that makes it to a global minimum I used 2.5 as the starting point.



# Linear Classification

## WU5

For each of the loss functions, train a model on the binary version of the wine data (called WineDataBinary) and evaluate it on the test data. You should use lambda=1 in all cases. Which works best? For that best model, look at the learned weights. Find the words corresponding to the weights with the greatest positive value and those with the greatest negative value. Hint: look at WineDataBinary.words to get the id-to-word mapping. List the top 5 positive and top 5 negative and explain.

The logistic loss function works best, with a training accuracy of 0.995951 and a test accuracy of 0.97417. In comparison, squared loss's training accuracy is 0.242915 and test accuracy is 0.313653, and hinge loss's training accuracy is 0.753036 and its test accuracy is 0.686347. All three classifiers used lambda = 1, numIter = 100, and stepSize = 0.5.
Top 5 positive:

1. citrus (0.8832897531176556)
2. crisp (0.7701247691557762)
3. lime (0.7108905521536929)
4. acidity (0.6891990079029967)
5. tropical (0.6064232619016873)

Top 5 negative:

1. tannins (-1.169521216404043)
2. black (-0.7653093906427071)
3. dark (-0.6835931677893784)
4. cherry (-0.6295907281434348)
5. blackberry (-0.5321916724675302)

The words corresponding with the most positive values, and therefore have the greatest impact on predicting the positive class are commonly associated with features of white wine. In contrast, the words corresponding with the most negative values are more associated with properties of red wine. As these two types of wines are polar opposites in texture, their corresponding words dividing the positive and negative classes does make sense.