WU1:

(A) For OAA, the binary decision tree related to classifying Sauvignon-Blanc is pasted below:

```
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-> opens?
|        |   -N-> class 1        (0 for class 0, 14 for class 1)
|        |   -Y-> class 0        (1 for class 0, 0 for class 1)
```

We see that when “lime” or “grapefruit” is present, the tree predicts Sauvignon-Blanc in the majority of cases, while if “flavors” is present, the tree predicts other wine most frequently. Therefore, “lime” and “grapefruit” are very indicative of being Sauvignon-Blanc, and “grapefruit” are indicative of being not.

Then we look at the AVA classifier. We print out the 4 trees associated with classifying Sauvignon-Blanc:

```
for i in range(1, 5):

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

The output is:

```
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)

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-> mild?
```

```
|        |  -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)

thai?
-N-> very?
|        -N-> friends?
|        |  -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)

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

We have the following observations:

- In the presence of “citrus” or “lime”, the first tree always predicts label1(Sauvignon-Blanc) against label0, accounting for 46 of the 60 label1 samples.
- Similarly, in the presence of “crispy” or “lime”, the second tree predicts label1 against label0 in most cases, accounting for 43 of the 60 label1 samples.
- In the third and fourth trees we don’t find a word that is very indicative of being Sauvignon-Blanc. As for words indicative of not being Sauvignon-Blanc, we notice that in the third tree the presence of “thai” always leads to not predicting label1, however, this accounts for only 5 samples.

Thus, in conclusion, the word “citrus”, “lime” and “crispy” are very indicative of being Sauvignon-Blanc. We find no words very indicative of not being Sauvignon-Blanc.

Using the same analysis for Pinot-Noir we are able to see:

For OAA, “raspberries” and “strawberry” are very indicative of being Pinot-Noir, while “cassis” and “petit” are very indicative of not being Pinot-Noir;

For AVA, “duck”, “cassis”, “acidity” (in the absence of “tannins”) are very indicative of being Pinot-Noir; “crisp”, “lime”, “lemon” and “straw” are very indicative of not being Pinot-Noir.

(B) Using AVA, training and prediction took 17.96 seconds in total; the accuracy is 0.2616; the most indicative words for Viognier are “peaches” and “milk”.

Using OAA, training and prediction took only 2.49 seconds; the accuracy is 0.3683; the most indicative words for Viognier are “floral”, “peach” and “peaches”.

(C) See the table below:

	Using zero-one	Using confidence
OAA	0.2468	0.3664
AVA	0.2672	0.2681

For OAA, using zero-one is markedly worse than using confidence, while for AVA the difference is negligible.

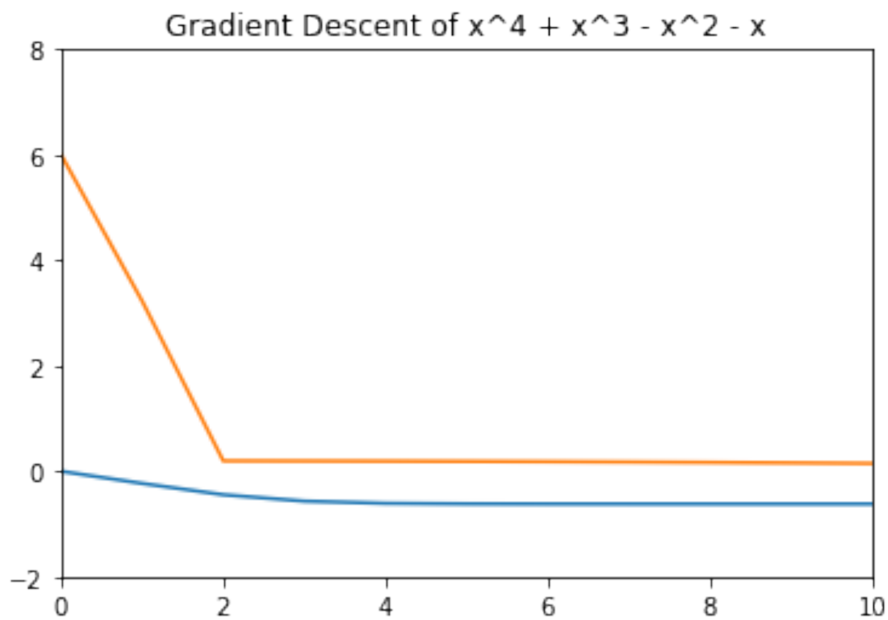
WU2 : The test accuracy is 0.3089.

WU3 (5%): 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^2 and 100 iterations, negative step sizes diverge. The step size of 0 is constant, while positive step sizes converge. Large positive step sizes either converge much later or simply diverge. As you reach the step size of 7, in 100 iterations the function no longer converges.

WU4 (10%): 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.)

The function f(x) = x^4 + x^3 - x^2 - x has a local minimum at x = -1 and a global minimum at x = 0.64. Using a start position of x = -2 will cause the function to get caught at the local minimum while using a start position of x = 0 will cause the function to converge to the global minimum.



WU5 (5%): 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.

```
LogisticLoss : Training accuracy 0.995951, test accuracy 0.97417
HingeLoss : Training accuracy 0.753036, test accuracy 0.686347
SquaredLoss : Training accuracy 0.242915, test accuracy 0.313653
```

The logistic loss works the best.

Top 5 positive:

- tropical
- acidity
- citrus
- lime
- crisp

Top 5 negative:

- blackberry
- cherry
- dark
- black
- tannins

Those 10 words' indexes map with the the weights' biggest absolute values. The reason might be that they represent significant nature of different wines, while other words are irrelevant to distinguish different wines.