Introduction to Machine Learning (CMSC-422)

Programming Assignment 1
Aalap Rana (116421393)

Multiclass and Linear Models

April 8, 2020

Contents

0.1 #WU 1

(A) Train depth 3 decision trees on the WineDataSmall task. What words are most indicative of being SauvignonBlanc Which words are most indicative of not being Sauvignon-Blanc? What about Pinot-Noir (label==2)?

If we consider OAA:

Label for wine Sauvignon-Blanc:

1. If we consider the word Citrus:
   Case Y : the word citrus is present:
   It was observed that 31 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 16 samples for label "Not Sauvignon-Blanc" by case 0.
   (citrus===65.95%(Sauvignon-Blanc) and 34.04%(Not Sauvignon-Blanc))
   Thus, presence of "Citrus" makes it likely to be "Sauvignon-Blanc".

2. If we consider the word grapefruit:
   Case Y : the word grapefruit is present:
   It was observed that 14 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 1 sample for label "Not Sauvignon-Blanc" by case 0.
   (grapefruit===93.33%(Sauvignon-Blanc) and 66.66%(Not Sauvignon-Blanc))
   Thus, presence of "grapefruit" makes it likely to be "Sauvignon-Blanc".

3. If we consider the word lime:
   Case Y : the word lime is present:
   It was observed that 15 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 3 samples for label "Not Sauvignon-Blanc" by case 0.
   (lime===83.33%(Sauvignon-Blanc) and 16.66%(Not Sauvignon-Blanc))
   Thus, presence of "lime" makes it likely to be "Sauvignon-Blanc".

4. If we consider the word or:
   Case Y : the word or is present:
   It was observed that 0 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 2 samples for label "Not Sauvignon-Blanc"

by case 0.
(or===0%(Sauvignon-Blanc) and 100%(Not Sauvignon-Blanc))
Thus, presence of "or" makes it likely to be "Not Sauvignon-Blanc".

5. If we consider the word gooseberry:
   Case Y : the word gooseberry is present:
   It was observed that 4 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 0 samples for label "Not Sauvignon-Blanc" by case 0.
   (gooseberry===100%(Sauvignon-Blanc) and 0%(Not Sauvignon-Blanc))
   Thus, presence of "gooseberry" makes it likely to be "Sauvignon-Blanc".

6. If we consider the word lingering:
   Case Y : the word lingering is present:
   It was observed that 0 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 1 sample for label "Not Sauvignon-Blanc" by case 0.
   (lingering===0%(Sauvignon-Blanc) and 100%(Not Sauvignon-Blanc))
   Thus, presence of "lingering" makes it likely to be "Not Sauvignon-Blanc".

7. If we consider the word flavour:
   Case Y : the word flavour is present:
   It was observed that 5 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 11 cases label "Not Sauvignon-Blanc" by case 0.
   (flavour===31.5%(Sauvignon-Blanc) and 68.75%(Not Sauvignon-Blanc))
   Thus, presence of "flavour" makes it likely to be "Not Sauvignon-Blanc".

Thus, the words whose presence makes the wine to be more likely to be of label "Sauvignon-Blanc "are :
Citrus
Gooseberry
Lime
Grapefruit

Thus, the words whose presence makes the wine to be more likely to be of label "Not Sauvignon-Blanc "are :
lingering
flavour
or

If we consider AVA:
Label for wine Sauvignon-Blanc:
If we consider The tree1:

1. If we consider the word Citrus:
   Case Y : the word citrus is present: It was observed that 31 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 0 samples for label "Not Sauvignon-Blanc" by case 0.
   (citrus===100%(Sauvignon-Blanc) and 0%(Not Sauvignon-Blanc))
   Thus, presence of "Citrus" makes it likely to be "Sauvignon-Blanc".

2. If we consider the word lime:
   Case Y : the word lime is present: It was observed that 15 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 0 samples for label "Not Sauvignon-Blanc" by case 0.
   (lime===100%(Sauvignon-Blanc) and 0%(Not Sauvignon-Blanc))
   Thus, presence of "lime" makes it likely to be "Sauvignon-Blanc".

3. If we consider the word melon:
   Case Y : the word melon is present: It was observed that 5 samples correseponds to label "Sauvignon-Blanc" as indicated by case 1 and 0 samples for label "Not Sauvignon-Blanc" by case 0.
   (melon===100%(Sauvignon-Blanc) and 0%(Not Sauvignon-Blanc))
   Thus, presence of "melon" makes it likely to be "Sauvignon-Blanc".

Thus, the words whose presence makes the wine to be more likely to be of label "Sauvignon-Blanc "are :
Citrus
Lime
melon

If we consider tree2:

1. If we consider the word Crisp:
   Case Y : the word crisp is present: It was observed that 30 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 2 samples for label "Not Sauvignon-Blanc" by case 0.
   (crisp===93.75%(Sauvignon-Blanc) and 6.25%(Not Sauvignon-Blanc))
   Thus, presence of "Crisp" makes it likely to be "Sauvignon-Blanc".

2. If we consider the word lime:
   Case Y : the word lime is present: It was observed that 13 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 1 samples for label "Not Sauvignon-Blanc" by case 0.
   (lime===92.85%(Sauvignon-Blanc) and 7.14%(Not Sauvignon-Blanc))
   Thus, presence of "lime" makes it likely to be "Sauvignon-Blanc".

3. If we consider the word lemon:
   Case Y : the word lemon is present: It was observed that 8 samples corresponds to label "Sauvignon-Blanc" as indicated by case 1 and 0 samples for label "Not Sauvignon-Blanc" by case 0.
   (lemon===100%(Sauvignon-Blanc) and 0%(Not Sauvignon-Blanc))
   Thus, presence of "lemon" makes it likely to be "Sauvignon-Blanc".

Thus, the words whose presence makes the wine to be more likely to be of label "Sauvignon-Blanc "are :
Crisp
Lime
lemon
If we do the same process for tree3 and tree4 and find the occurrence of words that predict label "Sauvignon-Blanc". We observe that the following words have maximum appearances and are likely to predict "Sauvignon-Blanc":
Crisp
Lime
lemon

The words indicative "Not Sauvignon-Blanc":
warm
pasta
red
apple

Label for wine Pinto-Noir:
If we consider OAA:

1. If we consider the word Cherry:
   Case Y : the word cherry is present: It was observed that 70 samples corresponds to label "Pinto-Noir" as indicated by case 1 and 65 samples label "Not Pinto-Noir " by case 0.
   (cherry===51.85%(Pinto-Noir) and 48.18%(Not Pinto-Noir))
   Thus, presence of "Cherry" makes it likely to be "Pinto-Noir".

2. If we consider the word Raspberries:
   Case Y : the word raspberries is present: It was observed that 12 samples correseponds to label "Pinto-Noir" as indicated by case 1 and 1 samples for label "Not Pinto-Noir" by case 0.
   (Raspberries===92.30%(Pinto-Noir) and 7.69%(Not Pinto-Noir))
   Thus, presence of "raspberries" makes it likely to be "Pinto-Noir".

3. If we consider the word Strawberry:
   Case Y : the word Strawberry is present: It was observed that 4 cases corresponds to label "Pinto-Noir" as indicated by case 1 and 0 cases label "Not Pinto-Noir" by case 0.
   (Strawberry===100%(Pinto-Noir) and 0%(Not Pinto-Noir))
   Thus, presence of "Strawberry" makes it likely to be "Pinto-Noir".

4. If we consider the word Full:
   Case Y : the word raspberries is present: It was observed that 0 cases corresponds to label "Pinto-Noir" as indicated by case 1 and 1 cases label "Not Pinto-Noir" by case 0.
   (Full===0%(Pinto-Noir) and 100%(Not Pinto-Noir))
   Thus, presence of "full" makes it likely to be "Not Pinto-Noir".

5. If we consider the word Cassis:
   Case Y : the word Cassis is present: It was observed that 2 cases corresponds to label "Pinto-Noir" as indicated by case 1 and 21 cases label "Not Pinto-Noir" by case 0.
   (Cassis===8.69%(Pinto-Noir) and 91.30%(Not Pinto-Noir))
   Thus, presence of "Cassis" makes it likely to be "Not Pinto-Noir".

6. If we consider the word Petit:
   Case Y : the word Petit is present: It was observed that 0 cases corre-seponded to label "Pinto-Noir" as indicated by case 1 and 8 cases label "Not Pinto-Noir" by case 0.
   (Petit===0%(Pinto-Noir) and 100%(Not Pinto-Noir))
   Thus, presence of "Petit" makes it likely to be "Not Pinto-Noir".

7. If we consider the word Allspice:
   Case Y : the word Allspice is present: It was observed that 2 samples corresponds to label "Pinto-Noir" as indicated by case 1 and 0 samples for label "Not Pinto-Noir" by case 0.
   (Allspices===100%(Pinto-Noir) and 0%(Not Pinto-Noir))
   Thus, presence of "Allspice" makes it likely to be "Pinto-Noir".

Thus, the words whose presence makes the wine to be more likely to be of label "Pinto-Noir" are :
Allspice
Strawberry
Raspberries
Cherry

Thus, the words whose presence makes the wine to be more likely to be of label "Not Pinto-Noir" are :
Full
Cassis
Petit

Label for wine Pinto-Noir:
If we consider AVA:
From Tree 0:
case 0 == Pinto-Noir
case 1 == Not Pinto-Noir

1. If we consider the word Crisp:
   Case Y : the word crisp is present: It was observed that 30 samples corresponds to label "Not Pinto-Noir" as indicated by case 1 and 2 samples for label "Pinto-Noir " by case 0.

(crisp===6.25%(Pinto-Noir) and 93.75%(Not Pinto-Noir))
Thus, presence of "Crisp" makes it likely to be "Not Pinto-Noir".

2. If we consider the word lime: Case Y : the word lime is present: It was observed that 13 samples corresponds to label "Not Pinto-Noir" as indicated by case 1 and 1 samples for label "Pinto-Noir " by case 0.
(lime===7.14%(Pinto-Noir) and 92.85%(Not Pinto-Noir))
Thus, presence of "lime" makes it likely to be "Not Pinto-Noir".

3. If we consider the word lemon:
Case Y : the word lemon is present: It was observed that 8 samples corresponds to label "Not Pinto-Noir" as indicated by case 1 and 0 samples for label "Pinto-Noir " by case 0.
(lemon===0%(Pinto-Noir) and 100%(Not Pinto-Noir))
Thus, presence of "lemon" makes it likely to be "Not Pinto-Noir".

4. If we consider the word mild:
Case Y : the word mild is present: It was observed that 0 samples corresponds to label "Not Pinto-Noir" as indicated by case 1 and 1 sample for label "Pinto-Noir " by case 0.
(mild===100%(Pinto-Noir) and 0%(Not Pinto-Noir))
Thus, presence of "mild" makes it likely to be "Pinto-Noir".

5. If we consider the word red:
Case Y : the word red is present: It was observed that 0 samples corresponds to label "Not Pinto-Noir" as indicated by case 1 and 2 samples label "Pinto-Noir " by case 0.
(red===100%(Pinto-Noir) and 0%(Not Pinto-Noir))
Thus, presence of "red" makes it likely to be "Pinto-Noir".

Thus from this tree 0 words indicative of Pinto-Noir:
mild
red
Similarly if we consider all other trees:
Tree 1: acidity salmon 100 tannins tea
Tree 3: red
Tree 4: red
From the above trees we can state the words predictive of label:

mild
red

(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?

The accuracy is 36.82% and the run time is 3.04sec for OAA on Wine Data.
The accuracy is 26.99% and the run time is 24.6sec for AVA on Wine Data.

If we consider OAA:
The words predictive of label "Viognier":

1. If we consider the word Peaches:
   Case Y : the word peaches is present: It was observed that 4 samples corresponds to label "Viognier" as indicated by case 1 and 14 samples for label "Not Viognier" by case 0.
   (peaches===22.22%(Viognier) and 77.77%(Not Viognier))
   Thus, presence of "peaches" makes it likely to be "Not Viognier".

2. If we consider the word Nectarine:
   Case Y : the word Nectarine is present: It was observed that 2 samples corresponds to label "Viognier" as indicated by case 1 and 13 samples for label "Not Viognier" by case 0.
   (Nectarine===13.33%(Viognier) and 86.66%(Not Viognier))
   Thus, presence of "Nectarine" makes it likely to be "Not Viognier".

3. If we consider the word Chilled:
   Case Y : the word Chilled is present: It was observed that 1 sample corresponds to label "Viognier" as indicated by case 1 and 6 samples for label "Not Viognier" by case 0.
   (Chilled===14.28%(Viognier) and 85.71%(Not Viognier)).
   Thus, presence of "Chilled" makes it likely to be "Not Viognier".

4. If we consider the word 5:
   Case Y : the word 5 is present: It was observed that 1 sample corresponds to label "Viognier" as indicated by case 1 and 0 samples for label "Not Viognier" by case 0.
   (5===100%(Viognier) and 0%(Not Viognier)).
   Thus, presence of "5" makes it likely to be "Viognier".

5. If we consider the word milk:
   Case Y : the word milk is present: It was observed that 3 samples corresponds to label "Viognier" as indicated by case 1 and 0 samples for label "Not Viognier" by case 0.
   (milk===100%(Viognier) and 0%(Not Viognier)).
   Thus, presence of "milk" makes it likely to be "Viognier".

6. If we consider the word fully:
   Case Y : the word fully is present: It was observed that 1 sample corresponds to label "Viognier" as indicated by case 1 and 0 samples for label "Not Viognier" by case 0.
   (fully===100%(Viognier) and 0%(Not Viognier)).
   Thus, presence of "fully" makes it likely to be "Viognier".

Thus the words indicative of Not Viognier are :
Peaches
Nectarine
Chilled
The words indicative of Viognier are:
Milk
5
fully

   If we consider AVA:
The words predictive of label "Viognier":
If we consider the Tree1:
class 0 === "Viognier" and class 1==="Not Viognier"

1. If we consider the word floral:
   Case Y : the word floral is present: It was observed that 1 sample corresponds to label " Not Viognier" as indicated by case 1 and 4 samples

for label "Viognier" by case 0.
(floral===80%(Viognier) and 20%(Not Viognier)).
Thus, presence of "floral" makes it likely to be "Viognier".

2. If we consider the word thai:
   Case Y : the word thai is present: It was observed that 0 sample corresponds to label " Not Viognier" as indicated by case 1 and 1 samples for label "Viognier" by case 0.
   (thai===100%(Viognier) and 0%(Not Viognier)).
   Thus, presence of "thai" makes it likely to be "Viognier".

3. If we consider the word enjoyed:
   Case Y : the word enjoyed is present: It was observed that 0 sample correseponds to label "Not Viognier" as indicated by case 1 and 1 samples for label "Viognier" by case 0.
   (enjoyed===100%(Viognier) and 0%(Not Viognier)) .
   Thus, presence of "enjoyed" makes it likely to be "Viognier".

4. If we consider the word juicy:
   Case Y : the word juicy is present: It was observed that 1 sample corresponds to label "Not Viognier" as indicated by case 1 and 0 samples for label "Viognier" by case 0.
   (juicy===0%(Viognier) and 100%(Not Viognier)).
   Thus, presence of "juicy" makes it likely to be "Not Viognier".

Thus from this tree 0 words indicating Vioginer:
floral
thai
enjoyed
A similar analysis for all the trees give the following list:
tree 1: peaches peach pear
tree 2 : peaches peach apple
tree 3 : peaches nectarine northern
tree 4 : peaches lovely also
Tree 5: milk also grenache
Tree 6: peaches apricot jasmine
Tree 7: floral pear straw
Tree 8: peach freshness apple

Tree 9: peaches peach apple
Tree 10: peaches peach apple
Tree 11: fruits peach pear
Tree 12: milk straw nuances
Tree 13: peaches peach pear
Tree 14: peach freshness apple
Tree 15: milk light lovely
Tree 16: floral peach straw
Tree 18: floral pear light
Tree 19: fruits white spice

Thus the word indicative of Vioginer:
peaches
peach
pear
Not indicative of Vioginer:
mineral

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

The accuracy for OAA on Wine Data Small is 58.86% for Zero one prediction and confidence prediction produces an accuracy of 60.61% . In case of Wine Data the accuracy is 24.76% for Zero one prediction and confidence prediction produces an accuracy of 36.27%.
The accuracy for AVA on Wine Data Small is 52.73% for Zero one prediction and confidence prediction produces an accuracy of 55.14% . In case of Wine Data the accuracy is 26.43% for Zero one prediction and confidence prediction produces an accuracy of 26.34%.
This is inline with the fact that OAA is sensitive to error, because binary classifiers are sensitive to error, and if any classifier makes error it can affect the overall vote count when the data is large with many classes. OAA has larger difference in Wine Data than in case of Wine Data Small. AVA has the least difference for both Wine Data and Wine Data Small.

0.2 #WU 2

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?

Implementing a tree-based reduction.The test accuracy of 30.89% was achieved with a balanced tree on the Wine Data using a Decision Tree Classifier with max depth 3.

0.3 #WU 3

What is the impact of the step size on convergence? Find values of the step size where the algorithm diverges and converges.

We know that step size places an important role in reaching the convergence to the global minima in gradient descent. If the step size is too small then for the fixed number of iteration we would not be able to reach the global minima, instead we would stop before it. For the function under consideration it was observed that:
The gradient decent algorithm does not converge to the global minima which in the given problem is 0. If we observe the last value of the trajectory it is 2.301773 at the 100th iteration for step size 0.05 and not equal to global minima of 0. It stops before the global minima.If we increase the number of iterations then it reaches the global minima 0.

If we make step size large. Then it converges to the global minima but experiences overshoots for medium values of step size here less than equal to 6 but any step size larger than 6 it doesnot reach the global minima. Thus there is divergence if the step size is too small or too large. Few cases were analysed :
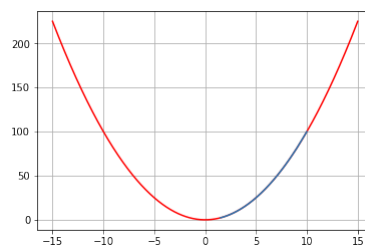Case 1: step size 0.05

Figure 1: Step Size Effect- Small value of 0.05
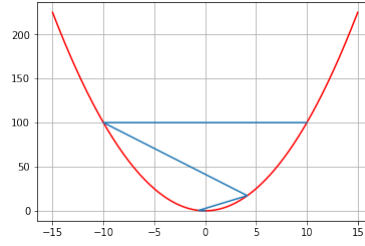
Case 2: step size 1

13

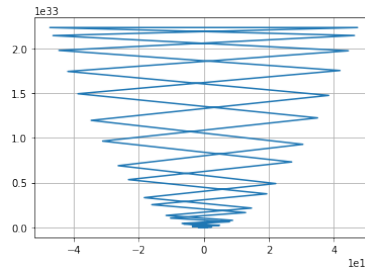Figure 2: Step Size Effect- Small value of 1

Case 3: step size 7

Figure 3: Step Size Effect- Large value of 7

0.4 #WU 4

Come up with a non-convex uni-variate optimization problem. Plot the function you're trying to minimize and show two runs of gd.py, one where it gets caught in a local minimum and one where it manages to make it to a global minimum?

The non-convex uni-variant function I have used is:

$$F(x) = sin(x)e^{-0.1*(x-0.6)^2} \tag{1}$$

The gradient of the function is:

$$F'(x) = (sin(x) * -0.2 * (x - 0.6) + cos(x))e^{-0.1*(x-0.6)^2} \tag{2}$$

We can observe form the graph of the function that the global minima is at -1.221 and a local minima is at 4.101.
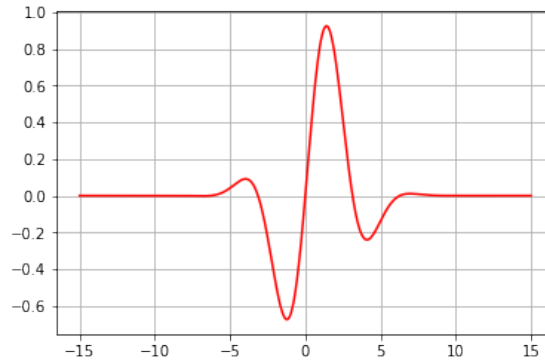
14

Figure 4: Non-convex Optimization Function

As we can observe that the gradient descent of the function attains the local minima when computation begins with positive values.We used x=2 and attained a value of 4.09292514539648 . It was observed that the gradient descent of the function attains the global minima when computation begins with negative values. We used x= -2 and attained a value of -1.2213994651276587. When the step size is too large it overshoots the global minima and attains the local minima. Here this happens for step size greater than or equal to 7.6 when the number of iteration is set constant at 100 and the start point is -2.

If the step size is too small the gradient descent function converges before reaching the golbal minima. When the number of iteration is set constant at 100 and the start point is -2. To over come the situation we have to increase the number of iteration.

0.5   #WU 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?

We observe that the dataset WineDataBinary gave the following accuracy for different Loss functions:

Squared Loss ——————Training accuracy 0.242915, test accuracy 0.313653
Logistic Loss ——————-Training accuracy 0.995951, test accuracy 0.97417

HingedLoss——————————Training accuracy 0.753036, test accuracy 0.686347

It can be observed that Logistic Loss performs best when we set the value of lambda to 1.

Words corresponding to the weights with the greatest positive value:

1)citrus

2)crisp

3)lime

4)acidity

5)tropical

Words corresponding to the weights with the greatest negative value:

1)tannins

2)black

3)dark

4)cherry

5)blackberry