CS 4210 – Assignment #2

### Maximum Points: 100 pts.

Bronco ID: |0|1|3|8|9|0|8|5|1|

Last Name: Manam

First Name: Viswadeep

**Note 1:** Your submission header must have the format as shown in the above-enclosed rounded rectangle.

**Note 2:** Homework is to be done individually. You may discuss the homework problems with your fellow students, but you are NOT allowed to copy – either in part or in whole – anyone else’s answers.

**Note 3:** Your deliverable should be a .pdf file submitted through Gradescope until the deadline. Do not forget to assign a page to each of your answers when making a submission. In addition, source code (.py files) should be added to an online repository (e.g., github) to be downloaded and executed later.

**Note 4:** All submitted materials must be legible. Figures/diagrams must have good quality.

**Note 5:** Please use and check the Canvas discussion for further instructions, questions, answers, and hints. The bold words/sentences provide information for a complete or accurate answer.

1. Timeline

   Description automatically generated with low confidence[16 points] Considering that ID3 built the decision tree below after analyzing a given training set, answer the following questions:
2. [12 points] What is the accuracy of this model if applied to the test set below? You must **identify** **each** True Positive, True Negative, False Positive, and False Negative for full credit.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | Spectacle | Astigmatism | Tear | Lenses (ground truth) | Result |
| Young | Hypermetrope | Yes | Normal | Yes | FP |
| Young | Hypermetrope | No | Normal | Yes | TP |
| Young | Myope | No | Reduced | No | TN |
| Presbyopic | Hypermetrope | No | Reduced | No | TN |
| Presbyopic | Myope | No | Normal | No | FN |
| Presbyopic | Myope | Yes | Reduced | No | TN |
| Prepresbyopic | Myope | Yes | Normal | Yes | TP |
| Prepresbyopic | Myope | No | Reduced | No | TN |

Accuracy = (TP + TN) / (FP + FN + TP + TN) = 6 / 8

1. [4 points] What is the precision, recall, and F1-measure of this model when applied to the same test set?

Precision = TP / (TP + FP) = 2 / 3

Recall = TP / (TP + FN) = 2 / 3

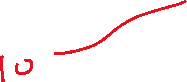
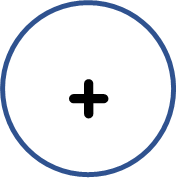
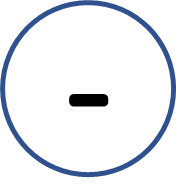
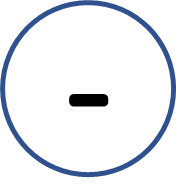
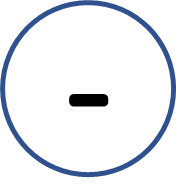
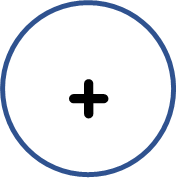
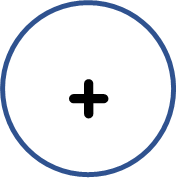
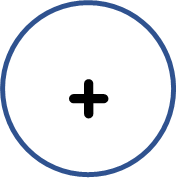
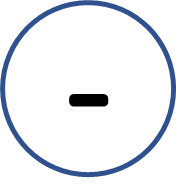
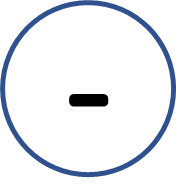
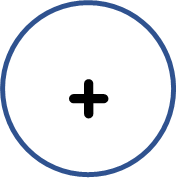
F1 = 2 \* ((2/3) \* (2/3)) / (2/3 + 2/3) = 2/3

1. [15 points] Complete the Python program (decision\_tree\_2.py) that will read the files contact\_lens\_training\_1.csv, contact\_lens\_training\_2.csv, and contact\_lens\_training\_3.csv. Each of those training sets has a different number of instances. You will observe that now the trees are being created setting the parameter *max\_depth = 3*, which it is used to define the maximum depth of the tree (pre-pruning strategy) in *sklearn*. Your goal is to train, test, and output the performance of the **3 models created by using each training set** on the test set provided (contact\_lens\_test.csv). **You must repeat this process 10 times** (train and test by using a different training set), choosing the lowest accuracy as the **final classification performance of each model**.

<https://github.com/SeveralCube22/CS4210_Assignments/tree/master/Assignment%202>

1. [32 points] Consider the dataset below to answer the following questions:

y



x

* 1. [4 points] What is the leave-one-out cross-validation error rate (LOO-CV) for **1NN**? Use Euclidean distance as your distance measure and the error rate calculated as:

If we leave out points #1, 2, 3, 4, we will make wrong predictions

Error rate = 4 / 10 = .4

* 1. [4 points] What is the leave-one-out cross-validation error rate (LOO-CV) for **3NN**?

If we leave out points #3, 4, 5, we will make wrong predictions

Error rate = 3 / 10 = .3

* 1. [4 points] What is the leave-one-out cross-validation error rate (LOO-CV) for **9NN**?

100%

* 1. [5 points] Draw de **decision boundary** learned by the 1NN algorithm.
  2. [15 points] Complete the Python program (knn.py) that will read the file binary\_points.csv and output the LOO-CV error rate for 1NN (**same answer of part a**).

<https://github.com/SeveralCube22/CS4210_Assignments/tree/master/Assignment%202>

1. [12 points] Find the class of instance #10 below following the 3NN strategy. Use Euclidean distance as your distance measure. You must **show all your calculations** for full credit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Red | Green | Blue | Class |
| #1 | 220 | 20 | 60 | 1 |
| #2 | 255 | 99 | 21 | 1 |
| #3 | 250 | 128 | 14 | 1 |
| #4 | 144 | 238 | 144 | 2 |
| #5 | 107 | 142 | 35 | 2 |
| #6 | 46 | 139 | 87 | 2 |
| #7 | 64 | 224 | 208 | 3 |
| #8 | 176 | 224 | 23 | 3 |
| #9 | 100 | 149 | 237 | 3 |
| #10 | 154 | 205 | 50 | ? |

Distance 1->10 =

Distance 2->10 = 149.26

Distance 3->10 =

Distance 4->10 = 100.12

Distance 5->10 = 80.02

Distance 6->10 = 131.87

Distance 7->10 = 182.83

Distance 8->10 = 39.67

Distance 9->10 = 202.53

Closest points = #4,5, 8

Majority class = 2

5. [25 points] Use the dataset below to answer the next questions:

Table

Description automatically generated

1. [10 points] Classify the instance ‹D15, Sunny, Mild, Normal, Weak› following the Naïve Bayes strategy. **Show all your calculations** until the final normalized probability values.

P(PlayTennis=NO | Outlook=Sunny, Temp=Mild, Humidity=Normal, Wind=Weak) =

P(Outlook=Sunny, Temp=Mild, Humidity=Normal, Wind=Weak |PlayTennis=NO) \* P(PlayTennis= NO)

Total = 14

NO = 5

YES = 9

P(Sunny | NO) = 3/5 = .6

P(Sunny | YES) = 2/9 = .22

P(Mild | NO) = 2 / 5 = .4

P(Mild | YES) = 4 / 9 = .44

P(Normal | NO) = 1 / 5 = .2

P(Normal | YES) = 6 / 9 =.66

P(Weak | NO) = 2 / 5 = .4

P(Weak | YES) = 6 / 9 = .66

P(NO) = .6 \* .4 \* .2 \* .4 \* (5/14) \* 100% = .686%

P(YES) = .22 \* .44 \* .66 \* .66 \* (9/14) \* 100% = 8.21%

1. [15 points] Complete the Python program (naïve\_bayes.py) that will read the file weather\_training.csv (training set) and output the classification of each test instance from the file weather\_test (test set) **if the classification confidence is >= 0.75**. Sample of output:

Day Outlook Temperature Humidity Wind PlayTennis Confidence

D15 Sunny Hot High Weak No 0.86

D16 Sunny Mild High Weak Yes 0.78

<https://github.com/SeveralCube22/CS4210_Assignments/tree/master/Assignment%202>

**Important Note:** Answers to all questions should be written clearly, concisely, and unmistakably delineated. You may resubmit multiple times until the deadline (the last submission will be considered).

**NO LATE ASSIGNMENTS WILL BE ACCEPTED. ALWAYS SUBMIT WHATEVER YOU HAVE COMPLETED FOR PARTIAL CREDIT BEFORE THE DEADLINE!**