

CS 4210 – Assignment #2

**Maximum Points: 100 pts.**

Bronco ID: | | | | | | | | | | Last Name: First Name:

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

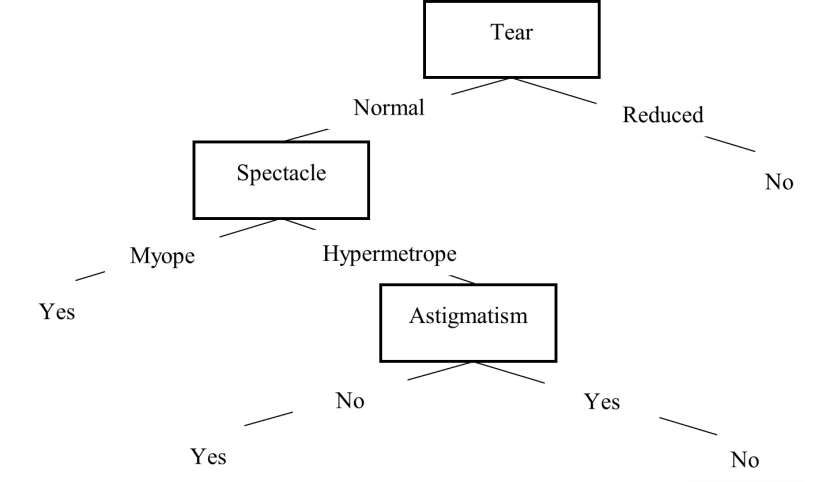


**Note 2:** Homework is to be done individually. You may discuss the homework problems with your fellow students, but you are NOT allowed to copy – either in part or in whole – anyone else’s answers.

**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. [16 points] Considering that ID3 built the decision tree below after analyzing a given training set, answer the following questions:
   1. [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. For instance:

TP = 1,5 | TN = 2,3 …

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Age | Spectacle | Astigmatism | Tear | Lenses (ground truth) |
| 1 | Young | Hypermetrope | Yes | Normal | Yes |
| 2 | Young | Hypermetrope | No | Normal | Yes |
| 3 | Young | Myope | No | Reduced | No |
| 4 | Presbyopic | Hypermetrope | No | Reduced | No |
| 5 | Presbyopic | Myope | No | Normal | No |
| 6 | Presbyopic | Myope | Yes | Reduced | No |
| 7 | Prepresbyopic | Myope | Yes | Normal | Yes |
| 8 | Prepresbyopic | Myope | No | Reduced | No |

|  |  |  |  |
| --- | --- | --- | --- |
| # | Decision Tree Output | Ground Truth | Identification |
| 1 | No | Yes | False Negative |
| 2 | Yes | Yes | True Positive |
| 3 | No | No | True Negative |
| 4 | No | No | True Negative |
| 5 | Yes | No | False Positive |
| 6 | No | No | True Negative |
| 7 | Yes | Yes | True Positive |
| 8 | No | No | True Negative |

True Positives: 2

True Negatives: 4

False Positives: 1

False Negatives: 1

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

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 average accuracy as the **final classification performance of each model**.
2. [32 points] Consider the dataset below to answer the following questions:

Icon  Description automatically generatedIcon  Description automatically generatedy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Icon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generatedIcon  Description automatically generated

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:

𝑛𝑢𝑚𝑏𝑒𝑟 𝑜𝑓 𝑤𝑟𝑜𝑛𝑔 𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑜𝑛𝑠

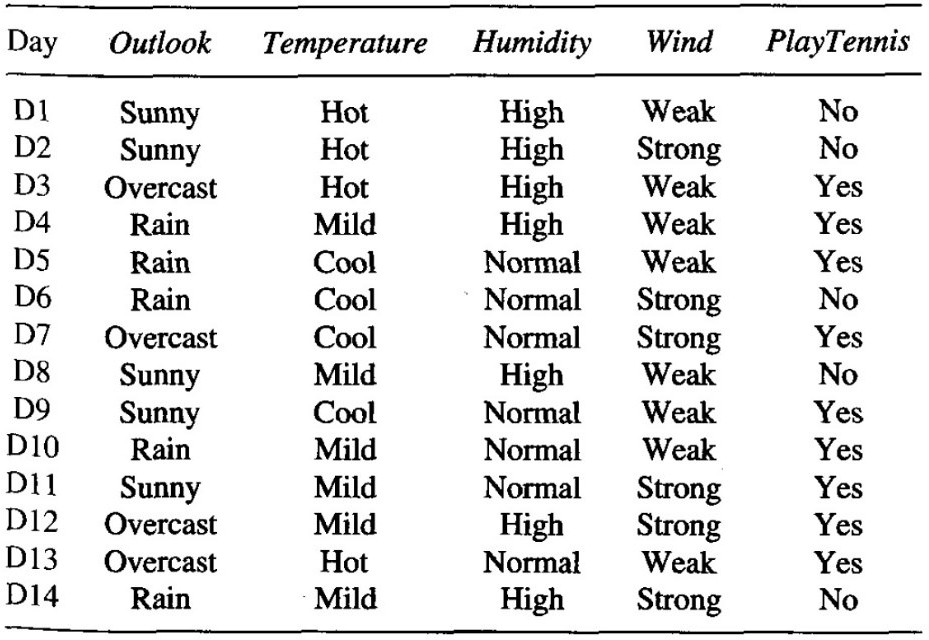
𝑒𝑟𝑟𝑜𝑟 𝑟𝑎𝑡𝑒 =

𝑡𝑜𝑡𝑎𝑙 𝑛𝑢𝑚𝑏𝑒𝑟 𝑜𝑓 𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑜𝑛𝑠

1. [4 points] What is the leave-one-out cross-validation error rate (LOO-CV) for **3NN**?
2. [4 points] What is the leave-one-out cross-validation error rate (LOO-CV) for **9NN**?
3. [5 points] Draw de **decision boundary** learned by the 1NN algorithm.
4. [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**).
5. [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 | ? |

1. [25 points] Use the dataset below to answer the next questions:



* 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.
  2. [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

**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!**