

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.

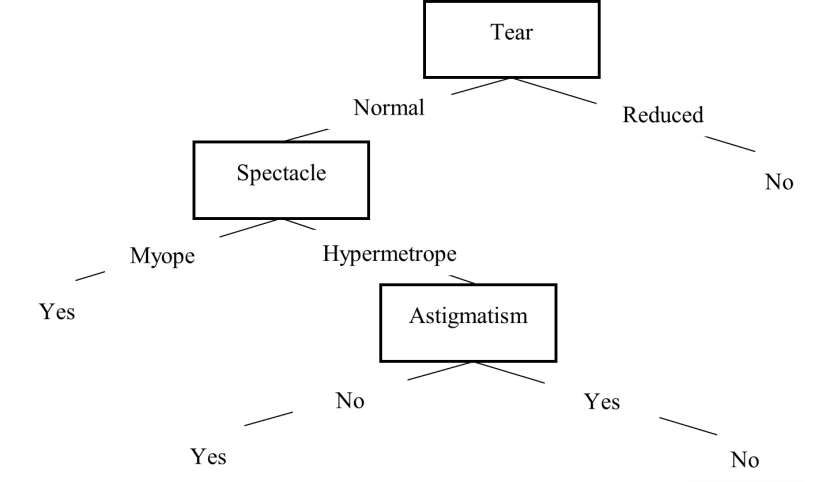


**Note 2:** Homework is to be done individually. You may discuss the homework problems with your fellow students, but you are NOT allowed to copy – either in part or in whole – anyone else’s answers.

**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, 7

True Negatives: 3, 4, 6, 8

False Positives: 5

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

|  |  |  |  |
| --- | --- | --- | --- |
| Number # | X | Y | Class |
| 1 | 0 | 5 | 0 |
| 2 | 0 | 3 | 0 |
| 3 | 1 | 4 | 0 |
| 4 | 2 | 4 | 1 |
| 5 | 2 | 1 | 0 |
| 6 | 3 | 3 | 1 |
| 7 | 3 | 2 | 1 |
| 8 | 4 | 4 | 1 |
| 9 | 4 | 3 | 1 |
| 10 | 4 | 1 | 0 |

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

Leave out point #1

The first nearest neighbor of point #1 is point #3 with distance 1.4142135623730951 and a class of negative

Correct Prediction

Leave out point #2

The first nearest neighbor of point #2 is point #3 with distance 1.4142135623730951 and a class of negative

Correct Prediction

Leave out point #3

The first nearest neighbor of point #3 is point #4 with distance 1.0 and a class of positive

Incorrect Prediction

Leave out point #4

The first nearest neighbor of point #4 is point #3 with distance 1.0 and a class of negative

Incorrect Prediction

Leave out point #5

The first nearest neighbor of point #5 is point #7 with distance 1.4142135623730951 and a class of positive

Incorrect Prediction

Leave out point #6

The first nearest neighbor of point #6 is point #7 with distance 1.0 and a class of positive

Correct Prediction

Leave out point #7

The first nearest neighbor of point #7 is point #6 with distance 1.0 and a class of positive

Correct Prediction

Leave out point #8

The first nearest neighbor of point #8 is point #9 with distance 1.0 and a class of positive

Correct Prediction

Leave out point #9

The first nearest neighbor of point #9 is point #6 with distance 1.0 and a class of positive

Correct Prediction

Leave out point #10

The first nearest neighbor of point #10 is point #7 with distance 1.4142135623730951 and a class of positive

Incorrect Prediction

Error Rate: 0.4

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

The 3 nearest neighbors to point #1 are:

Point #3 with a distance of 1.4142135623730951 and a class of 0

Point #2 with a distance of 2.0 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #1 is classified as negative

Correct Prediction

The 3 nearest neighbors to point #2 are:

Point #3 with a distance of 1.4142135623730951 and a class of 0

Point #1 with a distance of 2.0 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #2 is classified as negative

Correct Prediction

The 3 nearest neighbors to point #3 are:

Point #4 with a distance of 1.0 and a class of 1

Point #1 with a distance of 1.4142135623730951 and a class of 0

Point #2 with a distance of 1.4142135623730951 and a class of 0

Point #3 is classified as negative

Correct Prediction

The 3 nearest neighbors to point #4 are:

Point #3 with a distance of 1.0 and a class of 0

Point #6 with a distance of 1.4142135623730951 and a class of 1

Point #8 with a distance of 2.0 and a class of 1

Point #4 is classified as positive

Correct Prediction

The 3 nearest neighbors to point #5 are:

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #10 with a distance of 2.0 and a class of 0

Point #6 with a distance of 2.23606797749979 and a class of 1

Point #5 is classified as positive

Incorrect Prediction

The 3 nearest neighbors to point #6 are:

Point #7 with a distance of 1.0 and a class of 1

Point #9 with a distance of 1.0 and a class of 1

Point #4 with a distance of 1.4142135623730951 and a class of 1

Point #6 is classified as positive

Correct Prediction

The 3 nearest neighbors to point #7 are:

Point #6 with a distance of 1.0 and a class of 1

Point #5 with a distance of 1.4142135623730951 and a class of 0

Point #9 with a distance of 1.4142135623730951 and a class of 1

Point #7 is classified as positive

Correct Prediction

The 3 nearest neighbors to point #8 are:

Point #9 with a distance of 1.0 and a class of 1

Point #6 with a distance of 1.4142135623730951 and a class of 1

Point #4 with a distance of 2.0 and a class of 1

Point #8 is classified as positive

Correct Prediction

The 3 nearest neighbors to point #9 are:

Point #6 with a distance of 1.0 and a class of 1

Point #8 with a distance of 1.0 and a class of 1

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #9 is classified as positive

Correct Prediction

The 3 nearest neighbors to point #10 are:

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #5 with a distance of 2.0 and a class of 0

Point #9 with a distance of 2.0 and a class of 1

Point #10 is classified as positive

Incorrect Prediction

Error Rate: 0.2

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

The 9 nearest neighbors to point #1 are:

Point #3 with a distance of 1.4142135623730951 and a class of 0

Point #2 with a distance of 2.0 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #6 with a distance of 3.605551275463989 and a class of 1

Point #8 with a distance of 4.123105625617661 and a class of 1

Point #7 with a distance of 4.242640687119285 and a class of 1

Point #5 with a distance of 4.47213595499958 and a class of 0

Point #9 with a distance of 4.47213595499958 and a class of 1

Point #10 with a distance of 5.656854249492381 and a class of 0

Point #1 is classified as positive

Incorrect Prediction

The 9 nearest neighbors to point #2 are:

Point #3 with a distance of 1.4142135623730951 and a class of 0

Point #1 with a distance of 2.0 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #5 with a distance of 2.8284271247461903 and a class of 0

Point #6 with a distance of 3.0 and a class of 1

Point #7 with a distance of 3.1622776601683795 and a class of 1

Point #9 with a distance of 4.0 and a class of 1

Point #8 with a distance of 4.123105625617661 and a class of 1

Point #10 with a distance of 4.47213595499958 and a class of 0

Point #2 is classified as positive

Incorrect Prediction

The 9 nearest neighbors to point #3 are:

Point #4 with a distance of 1.0 and a class of 1

Point #1 with a distance of 1.4142135623730951 and a class of 0

Point #2 with a distance of 1.4142135623730951 and a class of 0

Point #6 with a distance of 2.23606797749979 and a class of 1

Point #7 with a distance of 2.8284271247461903 and a class of 1

Point #8 with a distance of 3.0 and a class of 1

Point #5 with a distance of 3.1622776601683795 and a class of 0

Point #9 with a distance of 3.1622776601683795 and a class of 1

Point #10 with a distance of 4.242640687119285 and a class of 0

Point #3 is classified as positive

Incorrect Prediction

The 9 nearest neighbors to point #4 are:

Point #3 with a distance of 1.0 and a class of 0

Point #6 with a distance of 1.4142135623730951 and a class of 1

Point #8 with a distance of 2.0 and a class of 1

Point #1 with a distance of 2.23606797749979 and a class of 0

Point #2 with a distance of 2.23606797749979 and a class of 0

Point #7 with a distance of 2.23606797749979 and a class of 1

Point #9 with a distance of 2.23606797749979 and a class of 1

Point #5 with a distance of 3.0 and a class of 0

Point #10 with a distance of 3.605551275463989 and a class of 0

Point #4 is classified as negative

Incorrect Prediction

The 9 nearest neighbors to point #5 are:

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #10 with a distance of 2.0 and a class of 0

Point #6 with a distance of 2.23606797749979 and a class of 1

Point #2 with a distance of 2.8284271247461903 and a class of 0

Point #9 with a distance of 2.8284271247461903 and a class of 1

Point #4 with a distance of 3.0 and a class of 1

Point #3 with a distance of 3.1622776601683795 and a class of 0

Point #8 with a distance of 3.605551275463989 and a class of 1

Point #1 with a distance of 4.47213595499958 and a class of 0

Point #5 is classified as positive

Incorrect Prediction

The 9 nearest neighbors to point #6 are:

Point #7 with a distance of 1.0 and a class of 1

Point #9 with a distance of 1.0 and a class of 1

Point #4 with a distance of 1.4142135623730951 and a class of 1

Point #8 with a distance of 1.4142135623730951 and a class of 1

Point #3 with a distance of 2.23606797749979 and a class of 0

Point #5 with a distance of 2.23606797749979 and a class of 0

Point #10 with a distance of 2.23606797749979 and a class of 0

Point #2 with a distance of 3.0 and a class of 0

Point #1 with a distance of 3.605551275463989 and a class of 0

Point #6 is classified as negative

Incorrect Prediction

The 9 nearest neighbors to point #7 are:

Point #6 with a distance of 1.0 and a class of 1

Point #5 with a distance of 1.4142135623730951 and a class of 0

Point #9 with a distance of 1.4142135623730951 and a class of 1

Point #10 with a distance of 1.4142135623730951 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #8 with a distance of 2.23606797749979 and a class of 1

Point #3 with a distance of 2.8284271247461903 and a class of 0

Point #2 with a distance of 3.1622776601683795 and a class of 0

Point #1 with a distance of 4.242640687119285 and a class of 0

Point #7 is classified as negative

Incorrect Prediction

The 9 nearest neighbors to point #8 are:

Point #9 with a distance of 1.0 and a class of 1

Point #6 with a distance of 1.4142135623730951 and a class of 1

Point #4 with a distance of 2.0 and a class of 1

Point #7 with a distance of 2.23606797749979 and a class of 1

Point #3 with a distance of 3.0 and a class of 0

Point #10 with a distance of 3.0 and a class of 0

Point #5 with a distance of 3.605551275463989 and a class of 0

Point #1 with a distance of 4.123105625617661 and a class of 0

Point #2 with a distance of 4.123105625617661 and a class of 0

Point #8 is classified as negative

Incorrect Prediction

The 9 nearest neighbors to point #9 are:

Point #6 with a distance of 1.0 and a class of 1

Point #8 with a distance of 1.0 and a class of 1

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #10 with a distance of 2.0 and a class of 0

Point #4 with a distance of 2.23606797749979 and a class of 1

Point #5 with a distance of 2.8284271247461903 and a class of 0

Point #3 with a distance of 3.1622776601683795 and a class of 0

Point #2 with a distance of 4.0 and a class of 0

Point #1 with a distance of 4.47213595499958 and a class of 0

Point #9 is classified as negative

Incorrect Prediction

The 9 nearest neighbors to point #10 are:

Point #7 with a distance of 1.4142135623730951 and a class of 1

Point #5 with a distance of 2.0 and a class of 0

Point #9 with a distance of 2.0 and a class of 1

Point #6 with a distance of 2.23606797749979 and a class of 1

Point #8 with a distance of 3.0 and a class of 1

Point #4 with a distance of 3.605551275463989 and a class of 1

Point #3 with a distance of 4.242640687119285 and a class of 0

Point #2 with a distance of 4.47213595499958 and a class of 0

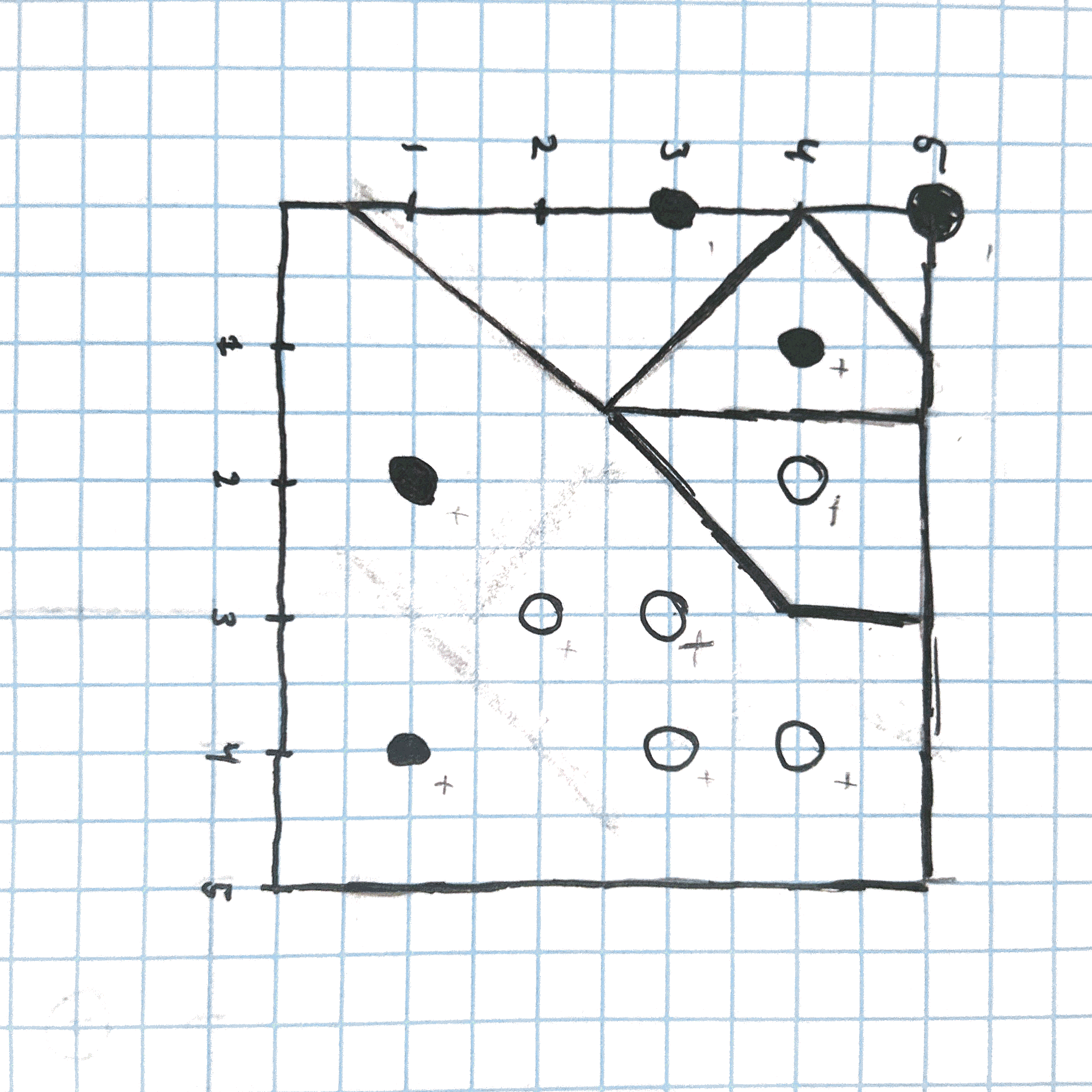
Point #1 with a distance of 5.656854249492381 and a class of 0

Point #10 is classified as positive

Incorrect Prediction

Error Rate: 1.0

1. [5 points] Draw the **decision boundary** learned by the 1NN algorithm.



1. [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/meap02/Assignment-2_CS4210-Machine-Learning-and-its-Applications>

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 | ? |

The 3 closest neighbors are:

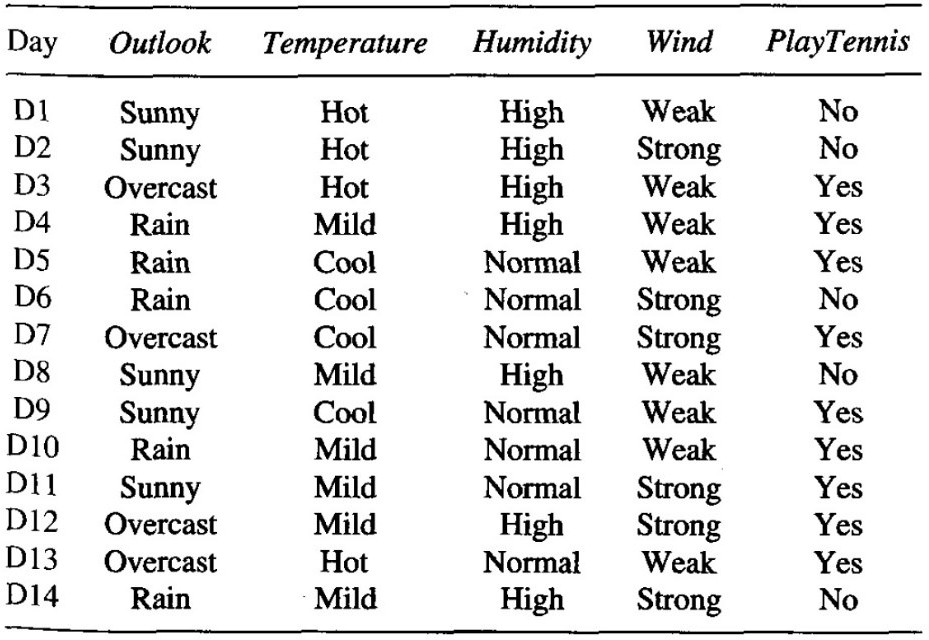
point #8 with distance 39.67366884975475 and class 3.0

point #5 with distance 80.01874780324921 and class 2.0

point #4 with distance 100.12492197250393 and class 2.0

This point is classified as class: 2.0

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.

So the most likely prediction class will be to go and play tennis.

* 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/meap02/Assignment-2_CS4210-Machine-Learning-and-its-Applications>

**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!**