Bronco ID: 015262624

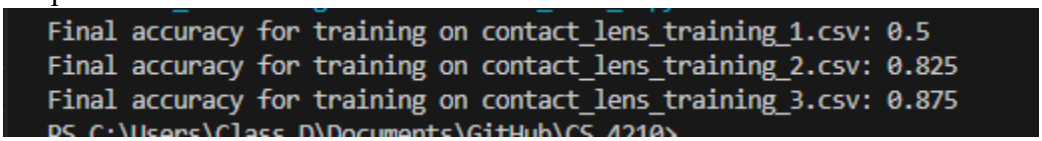
Last Name: Nguyen

First Name: Loc

1. Accuracy
    a. TP = 2, 7
       TN = 3, 4, 6, 8
       FP = 5
       FN = 1
    b. Precision = TP / (TP + FP) = 2 / (2+1) = 0.667 or 66.7%
       Recall = TP / (TP + FN) = 2 / (2 + 1) = 0.667 or 66.7%
       F1 = 2*0.667*0.667/ (0.667 + 0.667) = 0.667

2. Source Code – https://github.com/Skyhorizon2021/CS_4210/blob/main/Assignment2/decision_tree_2.py

   Output Screenshot



3a. 1NN Algorithm

Instance Coordinate	Predicted Class	True Class
2,1	+	-
4,1	+	-
3,2	+	+
0,3	-	-
3,3	+	+
4,3	+	+
1,4	+	-
2,4	-	+
4,4	+	+
0,5	-	-

Error rate = 4/10 = 40 %

3b. 3NN Algorithm

Instance Coordinate	Predicted Class	True Class
2,1	+	-
4,1	+	-
3,2	+	+
0,3	-	-
3,3	+	+
4,3	+	+
1,4	-	-
2,4	+	+
4,4	+	+
0,5	-	-

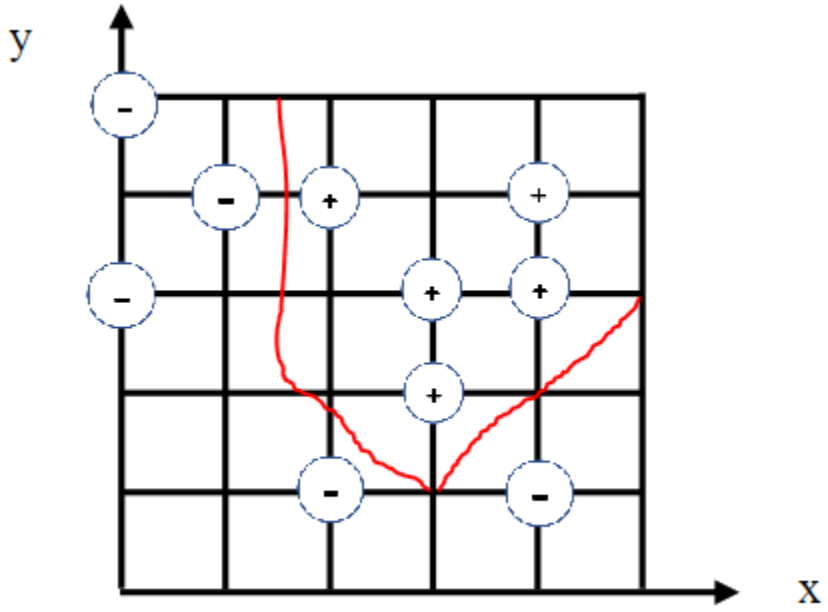
Error rate = 2/10 = 20%

3c. 9NN Algorithm

Instance Coordinate	Predicted Class	True Class
2,1	+	-
4,1	+	-
3,2	-	+
0,3	+	-
3,3	-	+
4,3	-	+
1,4	+	-
2,4	-	+
4,4	-	+
0,5	+	-

Error rate = 10 / 10 = 100%

3d. Decision boundary learned by 1NN algorithm



3e. Source Code - https://github.com/Skyhorizon2021/CS_4210/blob/main/Assignment2/knn.py

Output Screenshot

```
The error rate for 1NN algorithm: 0.4
```

4. 3NN Strategy

ID	Red	Green	Blue	Class	Distance to #10
1	220	20	60	1	196.67
2	255	99	21	1	274.35
3	250	128	14	1	281.21
4	144	238	144	2	313.23
5	107	142	35	2	181.21
6	46	139	87	2	170.31
7	64	224	208	3	312.31
8	176	224	23	3	285.80
9	100	149	237	3	297.27
10	154	205	50	2	0

Distance Calculations

$D(1,10) = \sqrt{(220 - 154)^2 + (20 - 205)^2 + (60 - 50)^2} = 196.67$

$D(2,10) = \sqrt{(225 - 154)^2 + (99 - 205)^2 + (21 - 50)^2} = 274.35$

$D(3,10) = \sqrt{(250 - 154)^2 + (128 - 205)^2 + (14 - 50)^2} = 281.21$

$D(4,10) = \sqrt{(144 - 154)^2 + (238 - 205)^2 + (144 - 50)^2} = 313.23$

$D(5,10) = \sqrt{(107 - 154)^2 + (142 - 205)^2 + (35 - 50)^2} = 181.21$

$D(6,10) = \sqrt{(46 - 154)^2 + (139 - 205)^2 + (87 - 50)^2} = 170.31$

$D(7,10) = \sqrt{(64 - 154)^2 + (224 - 205)^2 + (208 - 50)^2} = 312.31$

$D(8,10) = \sqrt{(176 - 154)^2 + (224 - 205)^2 + (23 - 50)^2} = 285.80$

$D(9,10) = \sqrt{(100 - 154)^2 + (149 - 205)^2 + (237 - 50)^2} = 297.27$

The 3 nearest neighbors' classes are one 1, and two 2. Therefore, using the 3NN strategy, class of instance #10 will be 2.

5a.

P (PlayTennis= No | (Outlook = Sunny, Temperature = Mild, Humidity = Normal, Wind = Weak):

P(Sunny|PlayTennis = No) * P(Mild|PlayTennis = No) * P(Normal|PlayTennis = No) * P(Weak|PlayTennis = No) * P(PlayTennis = No) => 3/5 * 2/5 * 1/5 * 2/5 * 5/14 = 0.007

P (PlayTennis= Yes | (Outlook = Sunny, Temperature = Mild, Humidity = Normal, Wind = Weak):

P(Sunny|PlayTennis = Yes) * P(Mild|PlayTennis = Yes) * P(Normal|PlayTennis = Yes) * P(Weak|PlayTennis = Yes) * P(PlayTennis = Yes) => 2/9 * 4/9 * 6/9 * 6/9 * 9/14 = 0.028

Normalization:

P (PlayTennis= No | (Outlook = Sunny, Temperature = Mild, Humidity = Normal, Wind = Weak)

- ⇨ 0.007 / (0.007 + 0.028) = 0.2

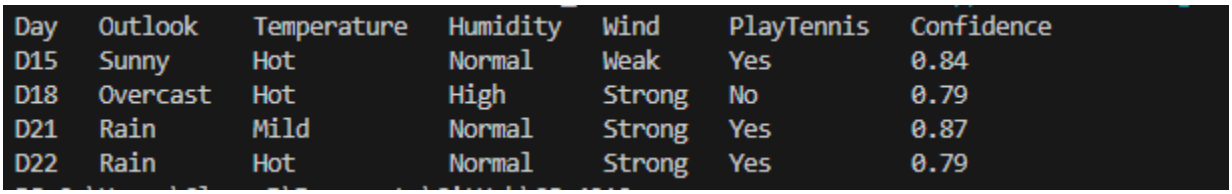
P (PlayTennis= Yes| (Outlook = Sunny, Temperature = Mild, Humidity = Normal, Wind =Weak):

- ⇨ 0.028 / (0.007 + 0.028) = 0.8

The most probable classification is Yes for PlayTennis

5b. Source Code - https://github.com/Skyhorizon2021/CS_4210/blob/main/Assignment2/naive_bayes.py

Screenshot Output



Day	Outlook	Temperature	Humidity	Wind	PlayTennis	Confidence
D15	Sunny	Hot	Normal	Weak	Yes	0.84
D18	Overcast	Hot	High	Strong	No	0.79
D21	Rain	Mild	Normal	Strong	Yes	0.87
D22	Rain	Hot	Normal	Strong	Yes	0.79