**HOMEWORK 8**

**Naive Bayes**

Assume we have the following data set that recorded (i.e., in a period of 25 days)

whether or not a person played tennis depending on the outlook and wind conditions.

Each instance (example) is represented by the three attributes.

Outlook: a value of {Sunny, Overcast, Rain}.

Wind: a value of {Weak, Strong}.

PlayTennis: the classification attribute (i.e., Yes means the person plays tennis; No means the person does not play tennis).

We want to predict if the person will play tennis in the three future days.

Day 26: (Outlook=Sunny, Wind=Strong) -> PlayTennis=?

Day 27: (Outlook=Overcast, Wind=Weak) -> PlayTennis=?

Day 28: (Outlook=Rain, Wind=Weak) -> PlayTennis=?

1. **PART-1: Manual Computation:--**Compute the predictions (i.e., the person will play tennis or not) for the three future days (Days 26-28) using the Naive Bayes classification approach.

**Day 26: (Outlook=Sunny, Wind=Strong) -> PlayTennis=?**

**SOLUTION: -**

**Step 1:**

P (Sunny | Yes), P (Strong | Yes)

P (Sunny | No), P (Strong | No)

**Step 2:**

Yes: No:

Sunny: Sunny:

n = 9 n = 9

n\_c = 5 n\_c = 4

p = 0.5 p = 0.5

m = 3 m = 3

Strong: Strong:

n = 10 n = 10

n\_c = 6 n\_c = 4

p = 0.5 p = 0.5

m = 3 m = 3

**Step 3:**

P (Sunny | Yes) = 5 + 3 \* 0.5 / 9 + 3 => **0.54**

P (Sunny | No) = 4 + 3 \* 0.5 / 9 + 3 => **0.46**

P (Strong | Yes) = 6 + 3 \* 0.5 / 10 + 3 => **0.58**

P (Strong | No) = 4 + 3 \* 0.5 / 10 + 3 => **0.42**

**Step 4:**

P (Yes) \* P (Sunny/Yes) \* P (Strong/Yes) => 0.5 \* 0.54 \* 0.58 => **0.156**

P (No) \* P (Sunny/ No) \* P (Strong/ No) => 0.5 \* 0.46 \* 0.42 => **0.096**

Since, 0.156 > 0.096, our example gets classified as **‘YES’**

**Day 27: (Outlook=Overcast, Wind=Weak) -> PlayTennis=?**

**SOLUTION: -**

**Step 1:**

P (Overcast | Yes), P (Weak | Yes)

P (Overcast | No), P (Weak | No)

**Step 2:**

Yes: No:

Overcast: Overcast:

n = 5 n = 3

n\_c = 5 n\_c = 3

p = 0.5 p = 0.5

m = 3 m = 3

Weak: Weak:

n = 15 n = 15

n\_c = 9 n\_c = 6

p = 0.5 p = 0.5

m = 3 m = 3

**Step 3:**

P (Overcast | Yes) = 5 + 3 \* 0.5 / 8 + 3 => **0.59**

P (Overcast | No) = 3 + 3 \* 0.5 / 8 + 3 => **0.41**

P (Weak | Yes) = 9 + 3 \* 0.5 / 15 + 3 => **0.58**

P (Weak | No) = 6 + 3 \* 0.5 / 15 + 3 => **0.42**

**Step 4:**

P (Yes) \* P (Overcast /Yes) \* P (Weak/Yes) => 0.5 \* 0.59 \* 0.58 => **0.171**

P (No) \* P (Overcast / No) \* P (Weak/ No) => 0.5 \* 0.41 \* 0.42 => **0.086**

Since, 0.171 > 0.086, our example gets classified as **‘YES’**

**Day 28: (Outlook=Rain, Wind=Weak) -> Play Tennis=?**

**SOLUTION: -**

**Step 1:**

P (Rain | Yes), P (Weak | Yes)

P (Rain | No), P (Weak | No)

**Step 2:**

Yes: No:

Rain: Rain:

n = 8 n = 8

n\_c = 5 n\_c = 3

p = 0.5 p = 0.5

m = 3 m = 3

Weak: Weak:

n = 15 n = 15

n\_c = 9 n\_c = 6

p = 0.5 p = 0.5

m = 3 m = 3

**Step 3:**

P (Rain | Yes) = 5 + 3 \* 0.5 / 8 + 3 => **0.59**

P (Rain | No) = 3 + 3 \* 0.5 / 8 + 3 => **0.41**

P (Weak | Yes) = 9 + 3 \* 0.5 / 15 + 3 => **0.58**

P (Weak | No) = 6 + 3 \* 0.5 / 15 + 3 => **0.42**

**Step 4:**

P (Yes) \* P (Rain/Yes) \* P (Weak/Yes) => 0.5 \* 0.59 \* 0.58 => **0.171**

P (No) \* P (Rain/ No) \* P (Weak/ No) => 0.5 \* 0.41 \* 0.42 => **0.086**

Since, 0.171 > 0.086, our example gets classified as **‘YES’**

1. **PART-2: Analysis with WEKA**

**SOLUTION: -**

**- How many instances used for the training? How many for the test?**

split 66.0% train, remainder test

**- How many instances are incorrectly classified?**

Incorrectly Classified Instances - 7

**- What is the MAE (mean absolute error) made by the classifier?**

Mean absolute error - 0.583

**- What can you infer from the information shown in the Confusion Matrix?**

=== Confusion Matrix ===

a b <-- classified as

1 1 | a = no

6 0 | b = yes

**- Visualize the classifier errors. In the plot, how can you differentiate between the correctly and incorrectly classified instances? In the plot, how can you see the detailed information of an incorrectly classified instance?**

By analysing the Detailed Accuracy By Class section

**- How can you save the learned classifier to a file?**

Right click on the “bayes.Naivebayes” which is present in result list, then by clicking on save result buffer

**- How many instances used for the training? How many for the test?**

user supplied test set: size unknown

**- How many instances are incorrectly classified?**

Incorrectly Classified Instances - 10

**- What is the MAE (mean absolute error) made by the classifier?**

Mean absolute error - 0.4719

**- What can you infer from the information shown in the Confusion Matrix?**

=== Confusion Matrix ===

a b <-- classified as

0 10 | a = no

0 18 | b = yes

**- Compare the test results with those observed in the previous experiment (i.e., using the splitting test mode).**

Splitting test has 12.5% correctly classified instances where as Supplied test has 64.3%.

Therefore, Result of Supplied test is better than splitting test option.