Confusion Matrix Understanding Multi-Class Machine Learning Model

Vikas Vyas

Calculate Precision and Recall for Multi-Class Classification Problem

The confusion matrix is used to visualize the performance of your prediction model.

With the help of the "sklearn.metrics import confusion_matrix" confusion matrix can be generated.

Agenda:

1. How 3*3 and 4*4 and so on dimensions are generated i.e n*m
2. How Precision and Recall calculated

Let's start with how dimensions are generated

Below given is the snapshot of the popular Iris dataset

```
1 irisDset.head()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

From the given dataset find out the unique Species. So, there are three unique Species Iris dataset contains.

```
1 irisDset.Species.unique()
```

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

Conclusion- Confusion matrix after completion of Model evaluation will be having a 3*3 dimensions. So, in case the dataset has 4 unique Species then matrix dimensions would be 4*4. Below given is one of the examples of the Iris dataset confusion matrix having 3 unique Species.

```
[[ 7  0  0]
 [ 0 10  2]
 [ 0  2  9]]
```

2. How Precision and Recall calculated

In order to calculate the Precision, Recall, and F1-Score, there is a need to find out the **TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative).**

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

All the diagonal values corresponding to each class are considered as TP (True Positive)

TP value for class Iris-setosa is 7

TP value for class Iris-versicolor is 10

TP value for class Iris-virginica is 9

The total number of False Positives for a class is the sum of the values of the corresponding column excluding (TP)

FP values for class Iris-setosa is (0+0) =0

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

FP values for class Iris-versicolor is (0 + 2)=2

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

FP value for class Iris-virginica is (0+2) = 2

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

The total number of False Negatives for a class is the sum of the values of the corresponding rows excluding (TP)

FN values for class Iris-setosa is (0+0) =0

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

FN values for class Iris-versicolor is (0 + 2)=2

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

FN value for class Iris-virginica is (0+2) = 2

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

The total number of True Negatives for a class is the sum of the values of all the columns and rows excluding that class's rows and columns

TN value for class Iris-setosa is (10+2+2+9)=23

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

TN value for class Iris-versicolor is (7+0+0+9)=16

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

TN value for class Iris-virginica is (7+0+0+10)= 17

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	7	0	0
Iris-versicolor	0	10	2
Iris-virginica	0	2	9

Calculate the Precision and Recall

Precision value for class Iris-setosa is TP/(TP+FP)=7/(7+0) =1

Precision value for class Iris-versicolor is TP/(TP+FP)=10/(10+2) =0.83

Precision value for class Iris-virginica is TP/(TP+FP)=9/(9+2) =0.82

Recall value for class Iris-setosa is TP/(TP+FN)=7/(7+0) =1

Recall value for class Iris-versicolor is TP/(TP+FN)=10/(10+2) =0.83

Recall value for class Iris-virginica is TP/(TP+FN)=9/(9+2) =0.82

All the above values can be directly calculated through

"from sklearn.metrics import classification_report"

Below given is classification report snapshot of popular Iris dataset

```
precision    recall

   1.00       1.00
   0.83       0.83
   0.82       0.82
```

precision   recall

   1.00   1.00