Abhishek Srivastava

Slot- E21+E22+E23

19BCE10071

Activity- 11

Que. Explore a classification problem case by considering any realworld domain application, formulate a confusion matrix through scenario assumption for the classifier model and investigate the various parameters to measure the performance of the classifier model.

Classification is the process of categorizing a given set of data into classes. In Machine Learning(ML), we frame the problem, collect and clean the data, add some necessary feature variables(if any), train the model, measure its performance, improve it by using some cost function, and then it is ready to deploy.
A great way to measure the performance of a classifier is through the use of Confusion Matrix.

→What is a Confusion Matrix?

A confusion matrix is a summary of prediction results on a classification problem.

The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix.

The confusion matrix shows the ways in which your classification model is confused when it makes predictions.

It gives you insight not only into the errors being made by your classifier but more importantly the types of errors that are being made.

It is this breakdown that overcomes the limitation of using classification accuracy alone.

→How to Calculate a Confusion Matrix

Below is the process for calculating a confusion Matrix.

1. You need a test dataset or a validation dataset with expected outcome values.
2. Make a prediction for each row in your test dataset.
3. From the expected outcomes and predictions count:
    1. The number of correct predictions for each class.
    2. The number of incorrect predictions for each class, organized by the class that was predicted.

- **Expected down the side**: Each row of the matrix corresponds to a predicted class.
- **Predicted across the top**: Each column of the matrix corresponds to an actual class. The counts of correct and incorrect classification are then filled into the table.

The total number of correct predictions for a class go into the expected row for that class value and the predicted column for that class value.

In the same way, the total number of incorrect predictions for a class go into the expected row for that class value and the predicted column for that class value.

This matrix can be used for 2-class problems where it is very easy to understand, but can easily be applied to problems with 3 or more class values, by adding more rows and columns to the confusion matrix.

➔ Example of 2-Class Confusion Matrix:-

Let's pretend we have a two-class classification problem of predicting whether a photograph contains a man or a woman.

We have a test dataset of 10 records with expected outcomes and a set of predictions from our classification algorithm.

1 Expected, Predicted

2 man, woman

3 man, man

4 woman, woman

5 man, man

6 woman, man
woman

7 woman,

8 woman, woman

9 man, man

10 man, woman
woman

11 woman,

Let’s start off and calculate the classification accuracy for this set of predictions.

The algorithm made 7 of the 10 predictions correct with an accuracy of 70%. accuracy

= total correct predictions / total

1 predictions made * 100
2
accuracy = 7 / 10 * 100

But what type of errors were made?

Let’s turn our results into a confusion matrix.

First, we must calculate the number of correct predictions for each class.

1 men classified as men: 3

2 women classified as women: 4

Now, we can calculate the number of incorrect predictions for each class, organized by the predicted value.

1 men classified as women: 2

2 woman classified as men: 1

We can now arrange these values into the 2-class confusion matrix:

1	men	women
2	men	3 1

3 women 2 4

We can learn a lot from this table.

- The total actual men in the dataset is the sum of the values on the men column (3 + 2) □ The total actual women in the dataset is the sum of values in the women column (1 +4).
- The correct values are organized in a diagonal line from top left to bottom-right of the matrix (3 + 4).
- More errors were made by predicting men as women than predicting women as men.