Siddhardhan

Accuracy Score & Confusion Matrix with Python implementation

Types of Supervised Learning

Supervised Learning

Classification

Regression

Classification is about predicting a
class or discrete values
Eg: Red or Green; True or False

Regression is about predicting a
quantity or continuous values
Eg: Salary; age; Price.

Evaluation metric for
Classification: **Accuracy score**

Evaluation metric for
Regression: **Mean Absolute Error**

SIDDHARDHAN

Accuracy Score

In Classification, **Accuracy Score** is the ratio of **number of correct predictions** to the **total number of input data points**.

$$\text{Accuracy Score} \ = \ \frac{\text{Number of correct predictions}}{\text{Total Number of data points}} \ \times \ 100 \ \%$$

Number of correct predictions = 128   Accuracy Score = 85.3 %

Total Number of data points = 150

```
from  sklearn.metrics  import  accuracy_score
```

SIDDHARDHAN

Limitation of Accuracy Score

Accuracy Score is not reliable when the dataset has an uneven distribution of classes

Number of dog images = 800   Number of cat images = 200

Number of images predicted as dog = 1000   Number of images predicted as cat = 0

Number of correct predictions = 800

Total Number of data points = 1000   Accuracy Score = $\dfrac{800}{1000}$ x 100 %

Accuracy Score = 80 %

Limitation of Accuracy Score

Accuracy Score is not reliable when the dataset has an uneven distribution of classes

Test data: Number of dog images   =   200   Number of cat images   =   200

Number of images predicted as dog = 400   Number of images predicted as cat = 0
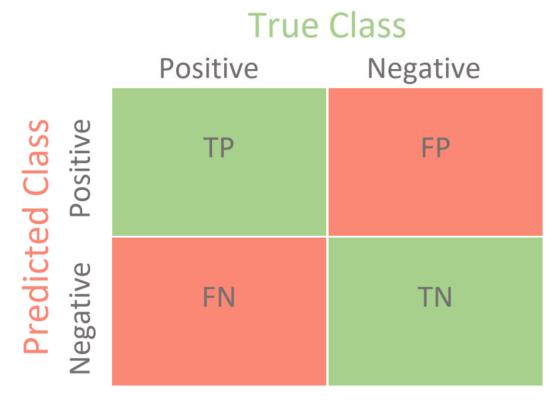
Number of correct predictions   =   200

Total Number of data points   =   400 $\text{Accuracy Score} = \dfrac{200}{400} \times 100 \ \%$

Accuracy Score   =   50 %

Confusion Matrix

Confusion Matrix is a matrix used for evaluating the performance of a Classification Model. It gives more information than the accuracy score.

TP + TN = Correct Predictions

FP + FN = Wrong Predictions

sklearn.metrics.confusion_matrix

SIDDHARDHAN