17.4.2

Confusion Matrix in Practice

Great job so far! Jill tells you that metrics such as sensitivity and precision can be a bit confusing at first. She assures you, however, that with enough practice, they will become second nature. She suggests that you return to a real-world dataset to deepen your understanding.

Let's look at an example of generating a confusion matrix in Python and interpreting it. To get started, download the following file:

Download 17-4-2-precision__recall_f1.zip (https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_17/17-4-2-precision__recall_f1.zip)

You now understand how precision, recall (sensitivity), and the F1 score can be used to assess a model's performance. Let's return to the Pima Indian diabetes dataset to go through an example in Python. Run all the cells in the notebook. All the data preparation steps have been performed, and a logistic regression model was trained and created predictions.

In the next cell, import the relevant modules for validation and print the `confusion_matrix`, which is the table of true positives, false positives, true negatives, and false negatives.

```
from sklearn.metrics import confusion_matrix, classification_report
matrix = confusion_matrix(y_test, y_pred)
print(matrix)
```

The table printed in the notebook is unlabeled, but can be interpreted as the following:

	Predicted True	Predicted False
Actually True	113	12

Actually False	31	36

How many true positives are there?

- ○ 113
- ○ 12
- ○ 31
- ○ 67

Check Answer

How many false positives are there?

- ○ 113
- ○ 12
- ○ 31
- ○ 67

Check Answer

What is the sensitivity/recall of this model?

- ○ 0.6
- ○ 0.7
- ○ 0.9
- ○ 1.0

Check Answer

What is the precision of this model, to two decimal places?

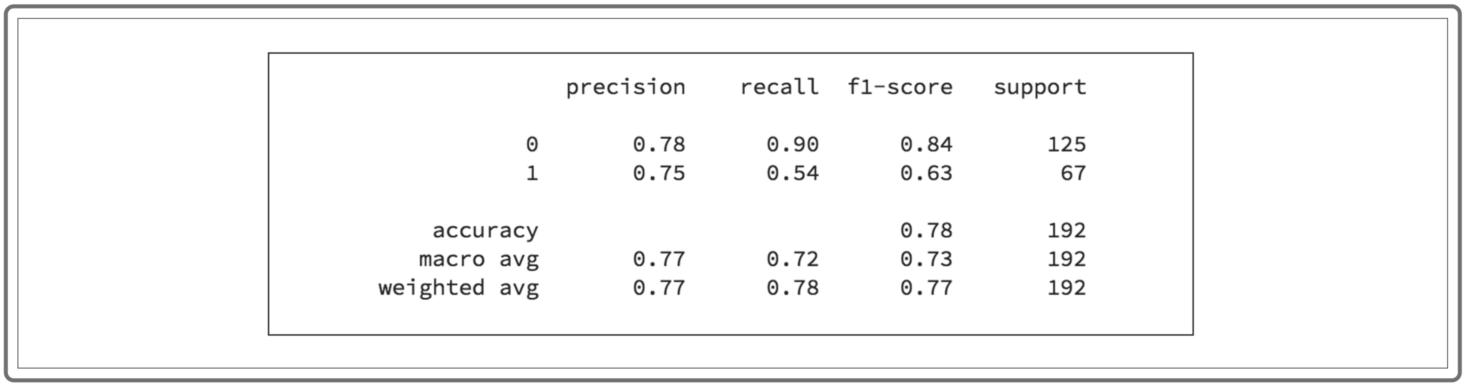
- ○ 0.6
- ○ 0.7
- ○ 0.78
- ○ 0.82

Check Answer

Finish ▶

Although we can manually calculate the metrics of the model, Scikit-learn's `classification_report` module performs the task for us:

```python
report = classification_report(y_test, y_pred)
print(report)
```



```
              precision    recall  f1-score   support

           0       0.78      0.90      0.84       125
           1       0.75      0.54      0.63        67

    accuracy                           0.78       192
   macro avg       0.77      0.72      0.73       192
weighted avg       0.77      0.78      0.77       192
```

The precision for prediction of the nondiabetics and diabetics are in line with each other. However, the recall (sensitivity) for predicting diabetes is much lower than it is for predicting an absence of diabetes. The lower recall for diabetics is reflected in the dropped F1 score as well.