100 XP

Evaluate classification models

4 minutes

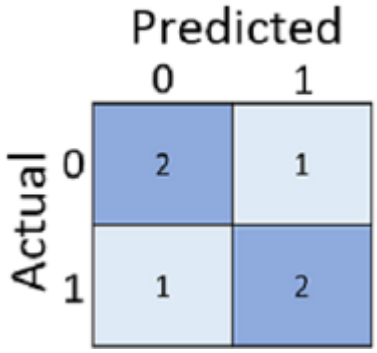
The training accuracy of a classification model is much less important than how well that model will work when given new, unseen data. After all, we train models so that they can be used on new data we find in the real world. So, after we have trained a classification model, we should evaluate how it performs on a set of new, unseen data.

In the previous units, we created a model that would predict whether a patient had diabetes or not based on their blood glucose level. Now, when applied to some data that wasn't part of the training set we get the following predictions:

x	y	ŷ
83	0	0
119	1	1
104	1	0
105	0	1
86	0	0
109	1	1

Recall that x refers to blood glucose level, y refers to whether they're actually diabetic, and ŷ refers to the model's prediction as to whether they're diabetic or not.

Simply calculating how many predictions were correct is sometimes misleading or too simplistic for us to understand the kinds of errors it will make in the real world. To get more detailed information, we can tabulate the results in a structure called a *confusion matrix*, like this:



The confusion matrix shows the total number of cases where:

- The model predicted 0 and the actual label is 0 (*true negatives*; top left)
- The model predicted 1 and the actual label is 1 (*true positives*; bottom right)
- The model predicted 0 and the actual label is 1 (*false negatives*; bottom left)
- The model predicted 1 and the actual label is 0 (*false positives*; top right)

The cells in the confusion matrix are often shaded so that higher values have a deeper shade. This makes it easier to see a strong diagonal trend from top-left to bottom-right, highlighting the cells where the predicted value and actual value are the same.

From these core values, you can calculate a range of other metrics that can help you evaluate the performance of the model. For example:

- **Accuracy**: (TP+TN)/(TP+TN+FP+FN) - out all of the predictions, how many were correct?
- **Recall**: TP/(TP+FN) - of all the cases that *are* positive, how many did the model identify?
- **Precision**: TP/(TP+FP) - of all the cases that the model predicted to be positive, how many actually *are* positive?

Next unit: Exercise - Perform classification with alternative metrics

Continue >