## How accurate is your model?

**Accuracy** is the answer to the question, "Out of all the patients, how many did we classify correctly?" The answer is the ratio between the number of correctly classified points and the number of total points. In this example, we can see that we have correctly classified 9,000 patients, which is a sum of 1,000 healthy and 8,000 sequences. So, the **accuracy** is this number divided by the total number of patients, which is 10,000. This gives us an accuracy of 90%.

Accuracy can easily be calculated as you can learn by using the accuracy score function.

|  |  |  |
| --- | --- | --- |
|  | Spam (positive) | Not Spam (negative) |
| Spam (positive) | True Positive(TP) | False Positive(FP) |
| Not Spam (negative) | False Negative(FN) | True Negative(TN) |

Often accuracy is not the only metric you should be optimizing on. This is especially the case when you have a class imbalance in your data. Optimizing on only accuracy can be misleading in how well your model is truly performing. With that in mind, you saw some additional metrics.

### Example: how much is the accuracy of the spam detector model?

This is the answer to the question, "Out of all the emails, how many did we classify correctly?" Since we have correctly classified 800 out of the 1,000 emails, then the answer is 80%.

### **Quiz Question**

Accuracy of 100% on a training set is like \_\_\_.

1. ....ignoring the training data. A model designed to have 100% accuracy on training data is likely to generalize well to new data.
2. ....memorizing the training data. A model designed to have 100% accuracy on training data is unlikely to generalize well to new data.
3. ....memorizing the training data. A model designed to have 100% accuracy on training data is likely to generalize well to new data.

### **Accuracy Quiz**

What is the accuracy of the model above? Please enter the answer as a percentage, with two decimals. For example, 54.75.

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Given the confusion matrix:

True Positives (TP) = 3

True Negatives (TN) = 2

False Positives (FP) = 1

False Negatives (FN) = 2

Accuracy Formula:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Calculation:

Accuracy = (3 + 2) / (3 + 2 + 1 + 2) = 5 / 8 = 0.625 (or 62.50%)

Final Answer:

62.50