

We can see that accuracy can be misleading if there are strongly imbalanced classes. In our cancer screening example, for instance, where only 1 person in 1,000 has cancer, a naive classifier that simply decides that nobody has cancer will achieve 99.9% accuracy and yet is completely useless.

Several other quantities can be defined in terms of these numbers, of which the most commonly encountered are

$$\text{Precision} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (5.30)$$

$$\text{Recall} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (5.31)$$

$$\text{False positive rate} = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TN}}} \quad (5.32)$$

$$\text{False discovery rate} = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TP}}} \quad (5.33)$$

In our cancer screening example, precision represents an estimate of the probability that a person who has a positive test does indeed have cancer, whereas recall is an estimate of the probability that a person who has cancer is correctly detected by the test. The false positive rate is an estimate of the probability that a person who is normal will be classified as having cancer, whereas the false discovery rate represents the fraction of those testing positive who do not in fact have cancer.

By altering the location of the decision boundary, we can change the trade-offs between the two kinds of errors. To understand this trade-off, we revisit [Figure 5.5](#), but now we label the various regions as shown in [Figure 5.10](#). We can relate the labelled regions to the various true and false rates as follows:

$$N_{\text{FP}}/N = E \quad (5.34)$$

$$N_{\text{TP}}/N = D + E \quad (5.35)$$

$$N_{\text{FN}}/N = B + C \quad (5.36)$$

$$N_{\text{TN}}/N = A + C \quad (5.37)$$

where we are implicitly considering the limit $N \rightarrow \infty$ so that we can relate number of observations to probabilities.

5.2.6 ROC curve

A probabilistic classifier will output a posterior probability, which can be converted to a decision by setting a threshold. As the value of the threshold is varied, we can reduce type 1 errors at the expense of increasing type 2 errors, or vice versa. To better understand this trade-off, it is useful to plot the *receiver operating characteristic* or *ROC curve* (Fawcett, 2006), a name that originates from procedures to measure the performance of radar receivers. This is a graph of true positive rate versus false positive rate, as shown in [Figure 5.11](#).

As the decision boundary in [Figure 5.10](#) is moved from $-\infty$ to ∞ , the ROC curve is traced out and can then be generated by plotting the cumulative fraction of