A 2×2 matrix is shown, but the same concept works for \( K \) classes, where a \( K \times K \) matrix would require \( K^2(K – 1) \) off-diagonal relative misclassification costs to be estimated and plugged in.]

In the credit scoring application of Table 13.1, analysts have estimated that a default is seven times worse than a miss. That is, giving credit to an applicant who will eventually default swallows the profit made from seven good customers. (In reality, risk is along a continuum, but to make it a two-class problem, default has to be defined to be something concrete like “more than 90 days late at least once in a 2-year period.”) The diagonals have no cost, as that is where the predictions are correct. Note that the costs can be actual amounts or can be normalized so the smallest cost is 1.0; it is only their relative magnitude that matters. (Note also, some software tools require negative numbers; some positive.) Now, the algorithm can find the minimum cost rather than the maximum PC solution. This will be much more useful.\(^5\)

The results can also be reported in a matrix, known as a confusion matrix, as shown in Table 13.2.

There, 10% of the data cases are defaults, but 19% are predicted as defaults. This is the result of defaults being more costly to miss, so there is a preference for false alarms over false dismissals. Overall, 15% are errors (3% false dismissals with a relative cost of 7 each, and 12% false alarms with a cost of 1 each), for a total cost of 0.33%. [Note that the overall cost is just the dot product of the cost and confusion matrices, and that it is a relative cost here, so only the relative (to total) population needs to be tracked in the confusion matrix.]

If the cost matrix (for a two-class problem) is hard to estimate and only a range of likely relative costs is known, you can refer to a Receiver Operating Characteristic (ROC) curve of

\(^5\) The model form, especially for a structurally unstable method like decision trees, is often very sensitive to small changes in the cost matrix. This is due to so many alternative models being explored by the data mining search method, and is normal, albeit unsettling.