

to be fraudulent. A sample is taken to develop a model, and oversampling is used to provide a balanced sample in light of the very low response rate. When applied to this sample ($n = 800$), the model ends up correctly classifying 310 frauds, and 270 nonfrauds. It missed 90 frauds, and classified 130 records incorrectly as frauds when they were not.

- Produce the confusion matrix for the sample as it stands.
- Find the adjusted misclassification rate (adjusting for the oversampling).
- What percentage of new records would you expect to be classified as fraudulent?

- 5.6** A firm that sells software services has been piloting a new product and has records of 500 customers who have either bought the services or decided not to. The target value is the estimated profit from each sale (excluding sales costs). The global mean is \$2128. However, the cost of the sales effort is not cheap—the company figures it comes to \$2500 for each of the 500 customers (whether they buy or not). The firm developed a predictive model in hopes of being able to identify the top spenders in the future. The lift and decile charts for the validation set are shown in Figure 5.13.

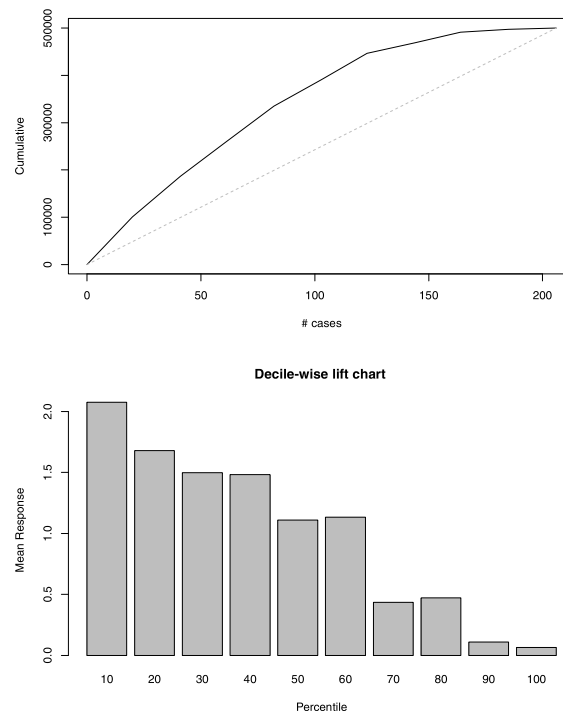


FIGURE 5.13 LIFT AND DECILE-WISE LIFT CHARTS FOR SOFTWARE SERVICES PRODUCT SALES

- If the company begins working with a new set of 1000 leads to sell the same services, similar to the 500 in the pilot study, without any use of predictive modeling to target sales efforts, what is the estimated profit?
- If the firm wants the average profit on each sale to at least double the sales effort cost, and applies an appropriate cutoff with this predictive model to a new set of 1000 leads, how far down the new list of 1000 should it proceed (how many deciles)?
- Still considering the new list of 1000 leads, if the company applies this predictive model with a lower cutoff of \$2500, how far should it proceed down the ranked leads, in terms of deciles?

- d. Why use this two-stage process for predicting sales—why not simply develop a model for predicting profit for the 1000 new leads?
- 5.7 Table 5.7 shows a small set of predictive model validation results for a classification model, with both actual values and propensities.
- a. Calculate error rates, sensitivity, and specificity using cutoffs of 0.25, 0.5, and 0.75.
- b. Create a decile-wise lift chart in R.

TABLE 5.7 PROPENSITIES AND ACTUAL CLASS MEMBERSHIP FOR VALIDATION DATA

Propensity of 1	Actual
0.03	0
0.52	0
0.38	0
0.82	1
0.33	0
0.42	0
0.55	1
0.59	0
0.09	0
0.21	0
0.43	0
0.04	0
0.08	0
0.13	0
0.01	0
0.79	1
0.42	0
0.29	0
0.08	0
0.02	0