Model Assessment Measures

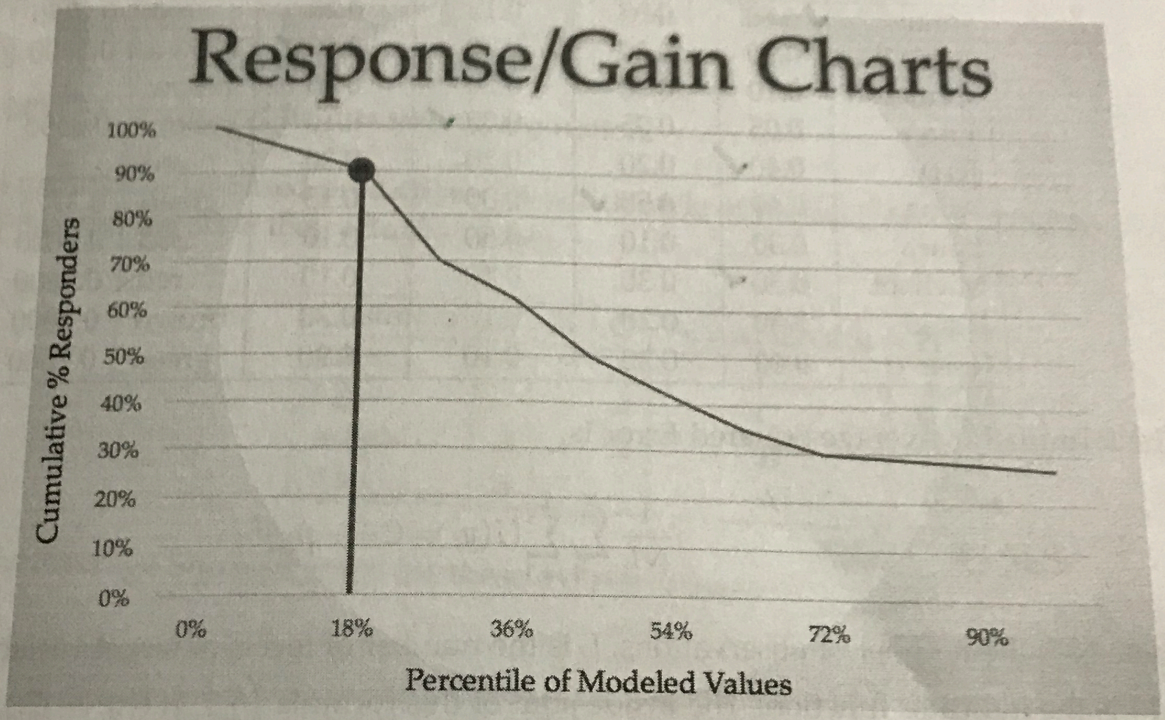
Exercises

1. If I'm most interested in ranking observations from most likely to least likely to respond, which model fit statistics might I find most useful in picking a model? (*Select all that apply.*)

   a. ROC index (AUC or c-statistic)
   b. Misclassification rate
   c. Average Squared Error
   d. Gini Coefficient
   e. KS-Statistic

   Potentially also these.

2. Answer the questions using the following plot of Cumulative % Responders vs. Model Depth:



a. Interpret the highlighted point in a sentence.

In the top ~20% of observations, as ranked by my model, we have a response rate of just over 90% (compared to ~27% in population)

b. What is the lift of the model at a depth of 18%?

~3
(just over 3)

3. Children pick M&M candies from four colors: red, blue, yellow, brown. We try to predict the probability p_{ij} that child i will pick color j based on the following variables:

- *gender*
- *age*
- *color picked by their friend*
- *claimed favorite color on the first day of school*
- *level of (hyper)activity on 1-10 scale*
- *color the teacher chooses*
- *shirt color that day*

For ten children, our model gives the estimated probabilities for choosing each color in the table below. For example, John has probability 0.6 of choosing red and his chosen color was actually red. Using this information, compute the quantities below.

Child	P_red	P_green	P_yellow	P_brown	Actual	SSQ
John	✓0.60	0.05	0.15	0.20	red	0.2250
Susie	0.10	0.30	0.20	0.40✓	brown	0.5000
Cameron	0.10	0.30	0.20	0.40	yellow	*a*
Elsa	0.05	0.05	0.80✓	0.10	yellow	0.0550
Jerry	0.40✓	0.20	0.20	0.20	red	*b*
Anna	0.10	0.50✓	0.30	0.15	green	0.3725
Laura	0.30	0.10	0.50	0.10	red	0.7600
Michael	0.50✓	0.30	0.10	0.10	red	0.3600
David	0.40	0.20	0.10	0.30	brown	0.7000
Jesse	0.40	0.30	0.10	0.20	green	0.7000

a. The formula for Average Squared Error is

$$\frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L}(I(y_i = C_j) - \hat{p}_{ij})^2$$

Where N is the number of observations, L is the number of levels of target variable $(C_j, j = 1, \ldots, L)$, and I is the indicator function. The inside sum of this formula $\left(\sum_{j=1}^{L}(I(y_i = C_j) - \hat{p}_{ij})^2\right)$ is given in the column SSQ. Fill in the missing values in this column:

$a = (0-0.1)^2 + (0-0.3)^2 + (1-0.2)^2 + (0-0.4)^2 = 0.01 + .09 + 0.64 + 0.16$

$a = 0.9$ $\quad b = 0.48$

$b = (1-0.4)^2 + (0-0.2)^2 + (0-0.2)^2 + (0-0.2)^2$
$= .36 + .04 \cdot 3$
$= .36 + .12$
$= .48$

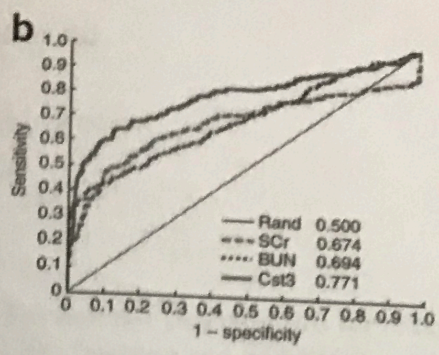
b. What is the Average Squared Error for this model?

0.1263

c. What is the misclassification rate for this model?

0.4

4. The following graph comes from the journal *Nature* and aims to compare 4 different models using a validation test. Based on this graph, which of the 4 models should the researchers conclude is performing best?



a.) Rand

b.) SCr

c.) BUN

d.) Cst3

e.) The ROC curve doesn't help answer this question.

5. The following are model-based predictions for which style of blouse (Style A, B, or C) a customer will choose, along with the style they actually chose:

Customer	A	B	C	Actual Choice	SSQ
John	.3	.3	.4	C	0.54
Lily	.1	.5	.4	C	0.62
Adam	.2	.6	.2	B	0.24
Alex	.8	.1	.1	A	0.96

Total: 0.365

Determine the Average Squared Error for these 4 observations.

$ASE = \frac{0.365}{4 \cdot 3} = 0.1216$

6. Why is it necessary to correct the modeled probabilities output from a model trained on over/undersampled data? because their accuracy relies on a representative sample, which we intentionally did not create.

7. Suppose you had a binary classification problem and you build two models. Think about what it would mean if you had two models with very similar ROC statistics, very similar misclassification rates, but one model had a substantially higher average squared error than the other. Can you imagine what the practical difference between these model outputs would be?