

DSC 433/533 Practice Mid-term Answer Key

1. **B** – old, traditional methods still have their place (e.g., old favorites multiple linear regression and logistic regression are often hard to beat).
2. **D** – lift is the ratio of cumulative responses/profit under the model (i.e., using model scores to order customers) to cumulative responses/profit without the model (i.e., using random selection).
3. **C** – see chapter 3 of the text-book.
4. **D** – R^2 in the training sample cannot be used to select a subset of predictors since it automatically increases (or stays the same) as you add a predictor (regardless of its value).
5. **A** – see response modeling exercise from class 5.
6. **D** – we need one fewer dummy indicator variables than there are categories, with one category representing a “reference level.”
7. **C** – better to use as much information as we can (this eliminates answers A and D) and using sensible numbers (C) is better than just using an arbitrary number like zero (B).
8. **C** – A is not a correct statement, and B and D, while correct statements, do not accurately describe the concept.
9. **D** – training error rates are irrelevant, and D has the (equal) lowest validation error rate and (equal) highest validation lift.
10. **A** – increasing the cut-off probability decreases the number of predicted purchasers (which rules out C and D), while B doesn’t keep the numbers of actual purchasers and non-purchasers fixed at 59 and 41 respectively.