

## Ch. 5 & 6 - Problem Bank Questions

September 13, 2020

1. Consider a subset selection procedure for linear regression. Answer the following questions:

- (a) Explain why using  $R^2$  to select models in either forward or backwards selection is not a good idea.
- (b) Forward and backwards selection are greedy approaches, in that they select the best variable at each step without looking ahead to what variables might be best overall. Explain why this can sometimes be an issue.
- (c) How does training MSE in general compare to the test MSE and why is this the case?
- (d) How do other metrics, such as  $C_p$ , AIC, BIC and adjusted  $R^2$  correct for the issues with  $R^2$  in model selections?

2. Consider figure 5.7 from the text (shown below). The purple dashed line is the Bayes optimal decision boundary.

(a) What can you say about the bias and variance of the logistic regression model with degree = 1?

(b) What can you say about the bias and variance of the logistic regression model with degree = 2?

(c) What can you say about the bias and variance of the logistic regression model with degrees = 3 and 4?

(d) Do you think that you should consider higher degrees for your logistic regression model based on these figures? Why or why not?

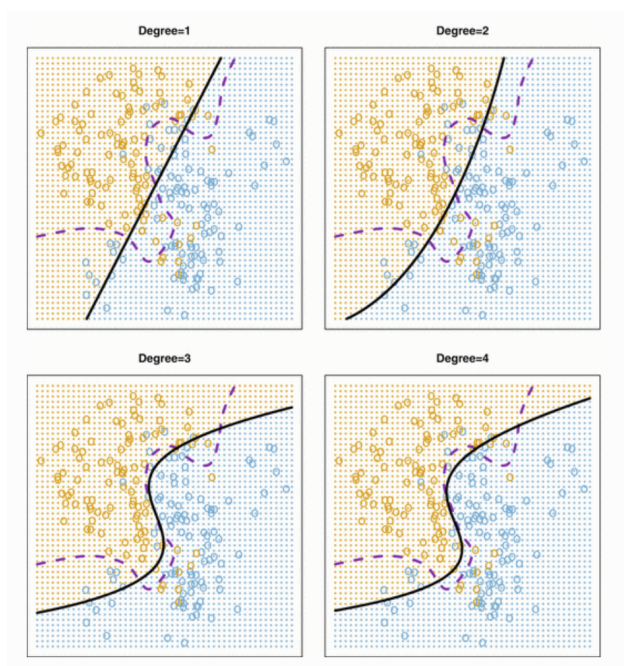


Figure 1: Figure 5.7 from the ISLR text.

4. Consider the following model-fitting scenarios. For each, describe how you would evaluate the accuracy or MSE of your model on some held-out test data. Does cross-validation make sense? If so, how many folds might you choose and why?

(a) A linear regression model fit to 10 data points.

(b) A logistic regression model fit to 5 million data points.

(c) A KNN classifier with  $K = 5$  that you can fit in 10 seconds.

(d) A complex neural network with 3 million parameters that takes one week to train.

5. Cross-validation is a commonly used method in machine learning.

(a) There are two main uses for cross-validation. What are they? Please give a specific example of each use case, and explain how they are different.

(b) Explain how k-fold cross-validation is implemented.

(c) What are the advantages and disadvantages of k-fold cross validation relative to:

i. The validation set approach?

ii. LOOCV?

(d). Higher degree polynomials could always achieve equivalent or better training error than lower order polynomials. Despite this, in graphs of LOOCV, we see times when a lower order polynomials has a lower MSE than a higher order polynomial. Why is this? Why is k-fold CV less likely to have this problem?

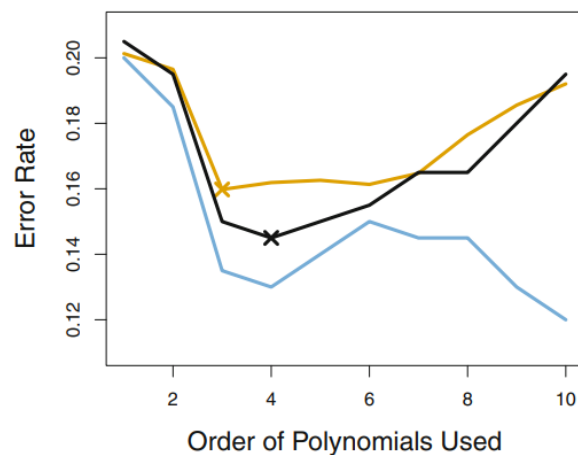


Figure 2: Figure 5.8 from the ISLR text.

6. AIC and BIC are two metrics that are commonly used for choosing between different models.

(a) Explain the conceptual difference between performing model selection with AIC/BIC vs. performing model selection with cross-validation. How do the processes differ?

(b) If  $n$  is the number of observations and  $d$  is the number of predictors in the model, AIC and BIC are defined as follows:

$$AIC = \frac{1}{n\sigma^2}(RSS + 2d\hat{\sigma}^2)$$
$$BIC = \frac{1}{n\sigma^2}(RSS + \log(n)d\hat{\sigma}^2)$$

Which metric has a higher penalty for including a larger number of predictors? Explain why this is the case, based on the formulas given above.

(c) Suppose you have a dataset and are picking between 10 models. You choose one model using AIC as the metric, and another model using BIC as the metric. Which of these models do you expect to have a lower variance (in terms of the bias-variance tradeoff)? Why?

8. ISL Chapter 6, Exercise 1