STA 325: Homework 2 (70 points)

DUE: 11:59pm, October 3 (on Sakai)
COVERAGE: ISL Chapters 5.1, 6.1

1. **[15 points]** We discussed in class two common ways of performing model selection. The first is using (i) a "test-error criterion" (e.g., AIC or BIC), and the second is using (ii) cross-validation.

   (a) Compare and contrast (i) and (ii) for model selection. What are the advantages and disadvantages of each type of method? When should a data analyst prefer one over the other?

   (b) For (i), compare and contrast AIC and BIC for model selection. When should a data analyst prefer one over the other?

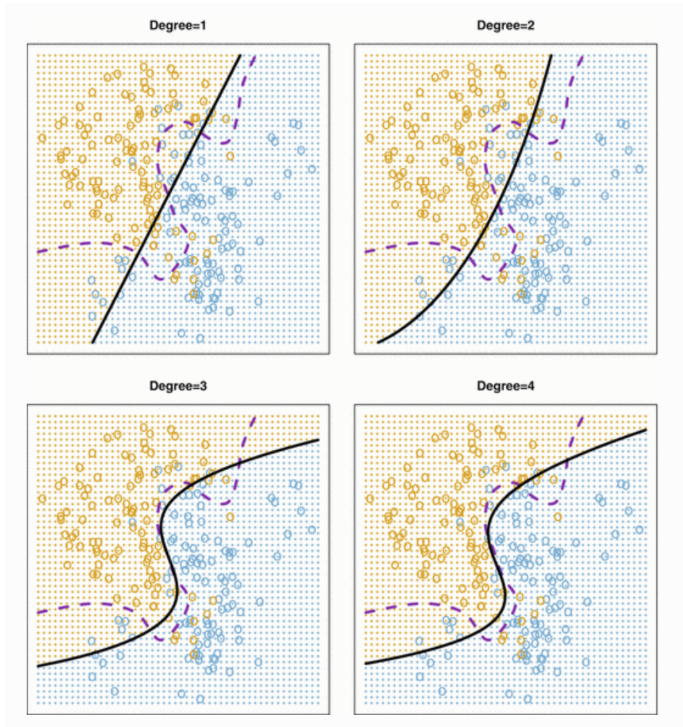
   (c) For linear regression, the AIC is equivalent to the classical Mallow's C_p criterion:

   $$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2).$$

   Intuitively, explain why the adjustment term $2d\hat{\sigma}^2$ should penalize models with more parameters d or larger (estimated) irreducible noise variance $\hat{\sigma}^2$.

   (d) Suppose you have a large dataset, and are choosing between 10 potential models. Let $\hat{f}_1$ be the fitted model selected by AIC, and $\hat{f}_2$ be the fitted model selected by BIC. Which model do you expect to have lower variance, $\text{Var}\{\hat{f}(x)\}$? Which model do you expect to have lower bias, $\text{Bias}\{\hat{f}(x)\}$? Explain your answer.

2. **[20 points]** Consider the following four classifiers, obtained by fitting logistic regression models with different polynomial degrees. The purple dotted lines show the Bayes-optimal classifier, and the black solid lines show the fitted classifier for the four logistic regression models.



More specifically, logistic regression is fit on the following four models:

$$\text{(degree 1):} \quad \operatorname{logit}\{p(x)\} = \beta_0 + \sum_{j=1}^{2} \beta_{j,1} x_j$$

$$\text{(degree 2):} \quad \operatorname{logit}\{p(x)\} = \beta_0 + \sum_{j=1}^{2} \beta_{j,1} x_j + \sum_{j=1}^{2} \beta_{j,2} x_j^2$$

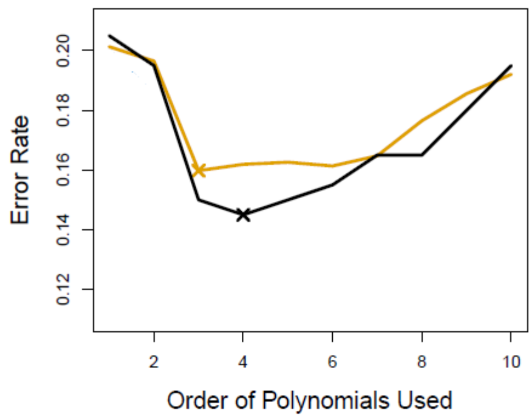
$$\text{(degree 3):} \quad \operatorname{logit}\{p(x)\} = \beta_0 + \sum_{j=1}^{2} \beta_{j,1} x_j + \sum_{j=1}^{2} \beta_{j,2} x_j^2 + \sum_{j=1}^{2} \beta_{j,3} x_j^3$$

$$\text{(degree 4):} \quad \operatorname{logit}\{p(x)\} = \beta_0 + \sum_{j=1}^{2} \beta_{j,1} x_j + \sum_{j=1}^{2} \beta_{j,2} x_j^2 + \sum_{j=1}^{2} \beta_{j,3} x_j^3 + \sum_{j=1}^{2} \beta_{j,4} x_j^4$$

(a) How many parameters are needed for each of the four models? How many parameters are needed for a polynomial model of degree M?

(b) What can you say about the bias and variance of the fitted model with degree = 1? With degree 2? With degrees 3 and 4?

(c) Plotted below are the true test errors (in black) and the estimated test errors using 10-fold CV (in orange) as a function of model polynomial degree. Comment on these error curves: What is the order of the true optimal model and the estimated optimal model? Why is the black curve (generally) lower than the orange curve?



(d) In this 0-1 classification problem, model complexity is varied by the order of polynomials used in logistic regression. Give two other ways to vary model complexity for this problem. For each, identify a variable which affects complexity (e.g., polynomial order), and explain how changing this variable affects the complexity of the fitted model.

3. **[20 points]** ISL Chapter 5, Exercise 8 (omit part f).

4. **[15 points]** ISL Chapter 6, Exercise 1.