MATH/STAT 4450/8456 Homework 2

Due date: February 7, 11:59 pm

1. (3 points) Derive the variance-bias decomposition for test MSE:

$$E(y_{new} - \hat{f}(x_{new}))^2 = \text{Var}\left(\hat{f}(x_{new})\right) + \left[\text{Bias}\left(\hat{f}(x_{new})\right)\right]^2 + \text{Var}(\epsilon)$$

2. (12 points) Textbook exercises.

   (a) (Chapter 2.4) 1

   (b) (Chapter 5.4) 2, 8

3. [Additional problem only for graduate students. Undergraduate students will receive 2 extra credits if solving the additional problems correctly.] (5 points)

   Suppose that a mechanism generates iid data pairs (x, y), according to the following model

$$x \sim Uniform(-\pi, \pi)$$

$$y|x \sim N\left(\sin(x), (0.5|x| + 0.5)^2\right)$$

   (a) Set the random seed in R by `set.seed(8456)`, then generate 200 data pairs use the mechanism above.

   (b) Use 4-fold cross validation on the generated data set to choose the best prediction model among the following six regressions using sets of predictors:
   $\{1, x\}$, $\{1, x, x^2\}$, $\{1, x, x^2, x^3\}$, $\{1, \sin x, \cos x\}$, $\{1, \sin x, \cos x, \sin 2x, \cos 2x\}$, and $\{1, x, x^2, x^3, \sin x, \cos x, \sin 2x, \cos 2x\}$.