Homework 1 (Sept. 21)

Deadline: Wednesday, October 5th, at 11:59pm.

Submission: You need to submit one file through Quercus with our answers to Questions 1, 2, 3, and 4, as well as R code and R outputs requested for Question 4. It should be a PDF file titled `hw1_writeup.pdf`. You can produce the file however you like (e.g. LaTeX, Microsoft Word, scanner), as long as it is readable.

Neatness Point: You will be deducted one point if we have a hard time reading your solutions or understanding the structure of your code.

Late Submission: 10% of the total possible marks will be deducted for each day late, up to a maximum of 3 days. After that, no submissions will be accepted.

- **Problem 1 (5 pts)**
  Assume that we have the regression model

$$Y = f(X) + \epsilon$$

  where ϵ is independent of X and $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[\epsilon^2] = \sigma^2$.

  1. (**2 pt**) Show that
$$f(X) = \mathbb{E}[Y \mid X].$$
  2. (**2 pts**) Prove that
$$\mathbb{E}[Y \mid X] = \underset{g}{\operatorname{argmin}}\, \mathbb{E}\Big[(Y - g(X))^2\Big].$$
     Hint: we have the fact that $\mathbb{E}[h(X,Y)] = \mathbb{E}_X \mathbb{E}_{Y|X}[h(X,Y) \mid X]$.
  3. (**1 pt**) Derive that
$$\mathbb{E}\Big[(Y - \mathbb{E}[Y \mid X])^2\Big] = \sigma^2.$$

  Parts (1) and (2) tell us that the best predictor of Y under the mean squared loss is $f(X)$. Part (3) points out why σ^2 is called the irreducible error.

- **Problem 2 (8 pts)**
  Assume that we have the regression model

$$Y = f(X) + \epsilon,$$

  where ϵ is independent of X and $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2$. Assume that the training data $(x_1, y_1), ..., (x_n, y_n)$ are used to construct an estimate of f, denoted by $\hat{f}$. Given a new random vector (X, Y) (independent of the training data),

1. (**3 pts**) show that

$$\mathbb{E}\Big[(f(x)-\hat{f}(x))^2 \mid X=x\Big] = \mathrm{Var}\Big(\hat{f}(x)\Big) + \Big[\mathbb{E}[\hat{f}(x)] - f(x)\Big]^2. \tag{0.1}$$

   Hint: You may benefit from adding and subtracting terms, such as

$$f(x)-\hat{f}(x) = f(x) - \mathbb{E}[\hat{f}(x)] + \mathbb{E}[\hat{f}(x)] - \hat{f}(x).$$

2. (**3 pts**) show that

$$\mathbb{E}\Big[\Big(Y-\hat{f}(x)\Big)^2 \mid X=x\Big] = \mathrm{Var}\Big(\hat{f}(x)\Big) + \Big(\mathbb{E}[\hat{f}(x)] - f(x)\Big)^2 + \sigma^2.$$

3. (**2 pt**) explain whether the expected MSE

$$\mathbb{E}\Big[\Big(Y-\hat{f}(X)\Big)^2\Big]$$

   can be smaller than σ^2 or not?

- **Problem 3 (8 pts)**

  Solve Problem 1 on page 52 (Chapter 2.4) in the textbook "Introduction to Statistical Learning". Each sub-question is worth 2 pts.

- **Problem 4 (16 pts)**

  This question should be answered using the Carseats data set which is contained in the R package ISLR. Each sub-question is worth 2 pts.

  (a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

  (b) Provide an interpretation of each coefficient in the model. Be careful–some of the variables in the model are qualitative!

  (c) Write out the model in equation form, being careful to handle the qualitative variables properly.

  (d) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$? Use the significance level 0.05 for the hypothesis test.

  (e) On the basis of your response to question (d), fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

  (f) What are the value of R^2 for models in (a) and (e)? Does larger R^2 mean the model fit the data better?

  (g) Using the model from (e), construct the 95 % confidence interval(s) for the coefficient(s).

  (h) Fit a linear regression model in (e) with interaction effect(s). Provide an interpretation of each coefficient in the model.