

Homework 1 (Sept. 20)

Deadline: Wednesday, October 4th, at 11:59pm.

Submission: You need to submit separate PDF files to each question via Crowdmark. For Question 5, your submission should also contain the R code and R outputs. You can produce the PDF's however you like (e.g. L^AT_EX, Microsoft Word, scanner), as long as they are legible.

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 (4 pts)** (This problem is to show you that why f is the best predictor of Y under the mean squared loss.)
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**) Prove that

$$f(X) = \mathbb{E}[Y | X].$$

2. (**2 pts**) Prove that

$$\mathbb{E}[Y | X] = \operatorname{argmin}_g \mathbb{E}[(Y - g(X))^2].$$

Hint: we have the fact that $\mathbb{E}[h(X, Y)] = \mathbb{E}_X \mathbb{E}_{Y|X}[h(X, Y) | X]$.

- **Problem 2 (6 pts)** (You will derive the Bias-Variance-Tradeoff formula in the lecture.)
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 $(x_1, y_1), \dots, (x_n, y_n)$ 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

$$\mathbb{E}\left[(f(x) - \hat{f}(x))^2 \mid X = x\right] = \text{Var}\left(\hat{f}(x)\right) + \left[\mathbb{E}[\hat{f}(x)] - f(x)\right]^2. \quad (0.1)$$

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

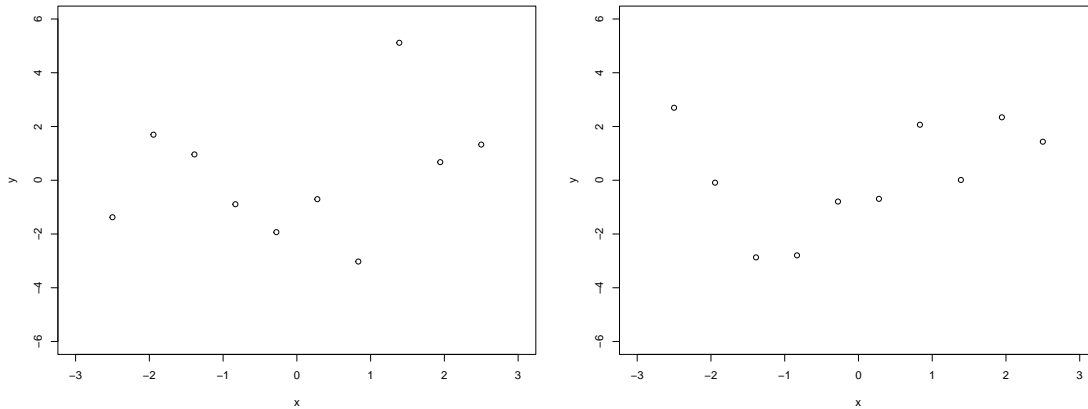
$$f(x) - \hat{f}(x) = f(x) - \mathbb{E}[\hat{f}(x)] + \mathbb{E}[\hat{f}(x)] - \hat{f}(x).$$

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

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

• **Problem 3 (6 pts)**

Suppose we have observed the following data points $(x_i, y_i)_{1 \leq i \leq 10}$ shown in the left panel. Further suppose we have observed a different set of data points generated from the same model (see, the right panel). Answer the following questions.



- (1 pt) Draw on both pictures what a fitted linear predictor (trained by using each data set) looks like (use the dotted line type or red).
- (1 pt) Draw on both pictures what an overfitted predictor (trained by using each data set) looks like (use the solid line type or black).
- (2 pts) For both the linear predictor and the overfitted predictor that you drew, state their predicted values for $x = 1.5$. (e.g. you should have two values for each type of predictor trained by using each data set)
- (2 pts) Comment on the predicted values in the previous part as well as the variances of these two types of predictors.

• Problem 4 (8 pts)

For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

- (a) (**2 pts**) The sample size n is extremely large, and the number of predictors p is small.
- (b) (**2 pts**) The number of predictors p is extremely large, and the number of observations n is small.
- (c) (**2 pts**) The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\epsilon)$, is extremely high.
- (d) (**2 pts**) The relationship between the predictors and response is highly non-linear.

• Problem 5 (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.