

# Statistical Learning

## Homework 2

### Part 1: Conceptual Questions

**Problem 1.** In an econometric model, we say that a parameter is identified if we can recover its value perfectly given the joint distribution of the observable variables. Suppose that  $(Y, X)$  is the observable variables and  $U$  is the unobservable variable.

1. Suppose that  $Y = \beta_0 + \beta_1 X + U$  and  $E[U] = E[XU] = 0$ . Show that  $\beta_1$  is identified. I.e., if you know the joint distribution of  $(Y, X)$ , how do you determine the value of the parameter  $\beta_1$ ?
2. Suppose that  $Y$  is binary and  $Y = 1(\beta_0 + \beta_1 X \geq U)$  and  $U$  is a standard normal  $N(0, 1)$  random variable that is independent of  $X$ . If you know the joint distribution of  $(Y, X)$ , how do you determine the value of the parameter  $\beta_1$ ? Hint:  $E[Y | X] = E[1(\beta_0 + \beta_1 X \geq U) | X] = \Phi(\beta_0 + \beta_1 X)$ , where  $\Phi$  is the standard normal CDF.

**Problem 2.** In this question, we show that in linear regression  $R^2$  is a non-decreasing function of the number of the regressors. Consider the sample  $(Y_i, X_{1,i}, X_{2,i})$ ,  $i = 1, 2, \dots, n$ , with two predictors  $X_{1,i}, X_{2,i}$ . Let  $\tilde{\beta}_0, \tilde{\beta}_1$  denote the OLS coefficients of the linear regression of  $Y_i$  against  $X_{1,i}$ . Let  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  denote the OLS coefficients of the linear regression of  $Y_i$  against  $X_{1,i}, X_{2,i}$ . Let  $\tilde{U}_i$  and  $\hat{U}_i$  denote the OLS residuals respectively. I.e.,

$$\begin{aligned} Y_i &= \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,i} + \tilde{U}_i, \\ Y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \hat{U}_i. \end{aligned}$$

1. Show that  $\sum_{i=1}^n \tilde{U}_i = \sum_{i=1}^n \tilde{U}_i X_{1,i} = 0$  and  $\sum_{i=1}^n \hat{U}_i = \sum_{i=1}^n \hat{U}_i X_{1,i} = \sum_{i=1}^n \hat{U}_i X_{2,i} = 0$ .
2. Show that  $\sum_{i=1}^n \tilde{U}_i \hat{U}_i = \sum_{i=1}^n \hat{U}_i^2$ .
3. Show that  $\sum_{i=1}^n \tilde{U}_i^2 \geq \sum_{i=1}^n \hat{U}_i^2$ .
4. Show that the  $R^2$  from the second (long) regression is larger than that of the first (short) regression.

**Problem 3.** Question 4 on Page 189 (ISL second edition).

**Problem 4.** Define a density function

$$f(x | \theta) = \begin{cases} \left(1 + \frac{1-2\theta}{\theta-1}\right) x^{\frac{1-2\theta}{\theta-1}} & x \in (0, 1) \\ 0 & x \notin (0, 1), \end{cases}$$

where  $0 < \theta < 1$  is a parameter.  $X_1, \dots, X_n$  is an independent and identically distributed sample with true density  $f(\cdot | \theta_*)$  for some  $\theta_*$ .

1. Show that  $f(\cdot | \theta)$  is a probability density function, for all  $0 < \theta < 1$ .
2. Show that  $\theta_* = \int_0^1 x f(x | \theta_*) dx$ . I.e., in this parametrization,  $\theta_*$  is also the population mean. Derive the method of moment estimator of  $\theta_*$ .
3. Write the log-maximum likelihood function and derive the maximum likelihood estimator. Is it equal to the method of moment estimator?

**Problem 5.** Given training data  $\text{Tr} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ , suppose that  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$  where  $\epsilon_i$  is the error term. Denote  $X_1^n = (X_1, \dots, X_n)$  for notational simplicity. Assume that  $E[\epsilon_i | X_1^n] = 0$ ,  $E[\epsilon_i^2 | X_1^n] = \sigma^2$  (for some  $\sigma^2 > 0$ ) and  $E[\epsilon_i \epsilon_j | X_1^n] = 0$ ,  $\forall i$  and  $\forall j \neq i$ . Assume that the conditional distribution of  $\epsilon_i$  given  $X_1^n$  is  $N(0, \sigma^2)$ . Let  $\hat{\beta}_0, \hat{\beta}_1$  denote the OLS estimator. Let  $x_0$  be a fixed value and  $y_0 = \beta_0 + \beta_1 x_0$ . Let  $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$  be the estimator of  $y_0$ . Let  $Y_0 = y_0 + \epsilon_0$ , where  $\epsilon_0$  denotes an error that is independent of the training data  $\text{Tr}$  ( $\epsilon_0 | \text{Tr} \sim N(0, \sigma^2)$ ). In this question, assume that  $\sigma^2$  is known.

1. Show that  $E[\hat{y}_0 | X_1^n] = y_0$  find the expression of  $\text{Var}[\hat{y}_0 | X_1^n]$ .
2. What is conditional distribution of  $\hat{y}_0$  given  $X_1^n$ ?
3. What is conditional variance of  $\hat{y}_0 - Y_0$  given  $X_1^n$ ? Hint:  $E[\epsilon_0 | \text{Tr}] = E[\epsilon_0] = 0$  and by law of iterated expectations,

$$E[\epsilon_0 \hat{y}_0 | X_1^n] = E[E[\epsilon_0 \hat{y}_0 | \text{Tr}]] = E[\hat{y}_0 E[\epsilon_0 | \text{Tr}]] = 0.$$

What is conditional distribution of  $\hat{y}_0 - Y_0$  given  $X_1^n$ ?

4. Propose a prediction interval  $[LB, UB]$  that covers  $Y_0$  with probability 95%. Find  $LB$  and  $UB$ .

## Part 2: Applied Questions

**Problem 6.** Question 8 on Page 123 (ISL second edition).

**Problem 7.** Question 9 on Page 123 (ISL second edition).

**Problem 8.** Question 13 on Page 193 (ISL second edition).

**Problem 9.** Question 14 on Page 194 (ISL second edition).