Statistical Learning

Homework 1

Part 1: Conceptual Questions

Problem 1. Let (X, Y) be a pair of random variables. Show that if $\mathrm{E}[Y \mid X] = \mathrm{E}[Y]$, then $\mathrm{Cov}[X, Y] = 0$.

Problem 2. Let (X, Y) be a pair of random variables. Denote $f(X) = \mathrm{E}[Y \mid X]$. Show that for any function g,

$$\mathrm{E}\left[(Y - f(X))^2 \mid X\right] \leq \mathrm{E}\left[(Y - g(X))^2 \mid X\right].$$

Hint: write

$$\mathrm{E}\left[(Y - g(X))^2 \mid X\right] = \mathrm{E}\left[(Y - f(X) + f(X) - g(X))^2 \mid X\right]$$

and use the law of iterated expectations (LIE).

Problem 3. Given training data $\mathsf{Tr} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ and a predictor $\hat{f}(x)$ which depends on Tr for any x, we have a new observation (X_0, Y_0) that is independent from Tr. Suppose that (X_0, Y_0) is generated by the model $Y_0 = f(X_0) + \epsilon_0$ with ϵ_0 being a new error term that is independent from (X_0, Tr). Show that the conditional expected test MSE can be decomposed into

$$\mathrm{E}\left[\left(Y_0 - \hat{f}(X_0)\right)^2 \mid X_0\right] = \mathrm{Var}[\epsilon] + \mathrm{Bias}(X_0)^2 + \mathrm{Variance}(X_0)$$

where $\mathrm{Bias}(X_0) = \mathrm{E}\left[\hat{f}(X_0) \mid X_0\right] - f(X_0)$ and

$$\mathrm{Variance}(X_0) = \mathrm{Var}\left[\hat{f}(X_0) \mid X_0\right] = \mathrm{E}\left[\left(\hat{f}(X_0) - \mathrm{E}\left[\hat{f}(X_0) \mid X_0\right]\right)^2 \mid X_0\right].$$

Hint: by LIE, write

$$\begin{aligned}
\mathrm{E}\left[\left(Y_0 - \hat{f}(X_0)\right)^2 \mid X_0\right] &= \mathrm{E}\left[\mathrm{E}\left[\left(Y_0 - \hat{f}(X_0)\right)^2 \mid X_0, \mathsf{Tr}\right] \mid X_0\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(Y_0 - f(X_0) + f(X_0) - \hat{f}(X_0)\right)^2 \mid X_0, \mathsf{Tr}\right] \mid X_0\right].
\end{aligned}$$

You may use the result $\mathrm{E}[Y_0 \mid \mathsf{Tr}, X_0] = \mathrm{E}[Y_0 \mid X_0]$ without proving it.

Problem 4. Suppose that Y is a binary response variable. The range of values taken by Y is $\{0,1\}$. The goal is to predict Y given another random variable X. When we observe a new X, we predict Y to be $h(X)$, where $h:\mathbb{R}\to\{0,1\}$ is a function that takes 0 or 1. We call h a classification rule. The "classification risk" of h is

$$R(h)=\Pr(Y\neq h(X)).$$

Let $m(x)=\mathrm{E}[Y\mid X=x]$. Since Y is binary,

$$\mathrm{E}[Y\mid X=x]=1\times\Pr(Y=1\mid X=x)+0\times\Pr(Y=0\mid X=x)=\Pr(Y=1\mid X=x).$$

(You may assume X is discrete if you have difficulty making sense of "$\Pr(Y=1\mid X=x)$". This is like $\Pr(A\mid B)$ with A being the event "$Y=1$" and B being the event $X=x$). Show that the rule that minimizes $R(h)$ is

$$h^*(x)=\begin{cases}1 & \text{if } m(x)>\frac{1}{2}\\ 0 & \text{otherwise.}\end{cases}$$

Hint: Note that

$$R(h)=\Pr(Y\neq h(X))=\int\Pr(Y\neq h(x)\mid X=x)\,f_X(x)\,dx,$$

where the second equality follows from LIE. It suffices to show that

$$\Pr(Y\neq h(x)\mid X=x)-\Pr(Y\neq h^*(x)\mid X=x)\geq 0 \text{ for all } x.$$

Use $\Pr(Y\neq h(x)\mid X=x)=1-\Pr(Y=h(x)\mid X=x)$ and

$$\Pr(Y=h(x)\mid X=x)=h(x)\Pr(Y=1\mid X=x)+(1-h(x))\Pr(Y=0\mid X=x).$$

Problem 5. Let $\{x_i:i=1,\ldots,n\}$ and $\{y_i:i=1,\ldots,n\}$ be two sequences. Define the averages

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n}x_i,\\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n}y_i.\end{aligned}$$

1. Show that $\sum_{i=1}^{n}(x_i-\bar{x})=0$.
2. Using the result in part (1), show that

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}x_i(x_i-\bar{x}),\ \text{and}\\ \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^{n}y_i(x_i-\bar{x})=\sum_{i=1}^{n}x_i(y_i-\bar{y}).\end{aligned}$$

Problem 6. Given training data $\mathsf{Tr} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$, suppose that $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where ϵ_i is the error term. The simple regression coefficient presented in class is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}.$$

Denote $X_1^n = (X_1, ..., X_n)$ for notational simplicity. Assume that $\mathrm{E}\left[\epsilon_i \mid X_1^n\right] = 0$, $\mathrm{E}\left[\epsilon_i^2 \mid X_1^n\right] = \sigma^2$ (for some $\sigma^2 > 0$) and $\mathrm{E}\left[\epsilon_i \epsilon_j \mid X_1^n\right] = 0$, $\forall i$ and $\forall j \neq i$.

1. Use the result in the last problem, show that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right) Y_i}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}.$$

2. Show that $\mathrm{E}\left[\hat{\beta}_1 \mid X_1^n\right] = \beta_1$ and $\mathrm{Var}\left[\hat{\beta}_1 \mid X_1^n\right] = \sigma^2 / \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$, where $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$.

3. Assume that the conditional distribution of ϵ_i given X_1^n is $\mathrm{N}\left(0, \sigma^2\right)$. What is the conditional distribution of Y_i given X_1^n?

4. What is the conditional distribution of $\hat{\beta}_1$ given X_1^n?

5. What is the unconditional distribution of the z-statistic:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}}?$$

Problem 7. ISL (2nd edition) Question 7 on Page 54.

Part 2: Applied Questions

Write your answer in an RMarkdown report, print your report and hand in.

Problem 8. ISL (2nd edition) Question 8. Give answers to Parts a, b and c(i-iv).

Problem 9. ISL (2nd edition) Question 9. Give answers to Parts a, b, c and d.