Solutions for Homework 1

Tong Zhou

Problems: 2.4, 2.5, 2.16, 2.19.

PROBLEM 2.4

Proof. Since both X and Y are discrete, just by definition, we have

$$\begin{aligned}
\mathbb{E}[Y \mid X = 0] &= \sum_{y \in \{0,1\}} y \cdot \mathbb{P}(Y = y \mid X = 0) = 1 \times \frac{0.4}{0.5} = 0.8 \\
\mathbb{E}[Y^2 \mid X = 0] &= \sum_{y \in \{0,1\}} y^2 \cdot \mathbb{P}(Y = y^2 \mid X = 0) = 0.8 \\
\mathsf{Var}(Y \mid X = 0) &= \mathbb{E}[Y^2 \mid X = 0] - (\mathbb{E}[Y \mid X = 0])^2 = 0.16.
\end{aligned}$$

Similarly when conditional on $X = 1$,

$$\begin{aligned}
\mathbb{E}[Y \mid X = 1] &= 0.6 \\
\mathbb{E}[Y^2 \mid X = 1] &= 0.6 \\
\mathsf{Var}(Y \mid X = 1) &= \mathbb{E}[Y^2 \mid X = 1] - (\mathbb{E}[Y \mid X = 1])^2 = 0.24.
\end{aligned}$$

□

PROBLEM 2.5

Proof. (a) By definition the mean-squared error of e^2 given X is:

$$\mathbb{E}\left[e^2 - h(X)\right]^2.$$

(b) To predict e^2, one seeks to find a function h such that the MSE is minimized.
(c) Project e^2 on the space of X, we have

$$e^2 = \mathbb{E}\left[e^2 \mid X\right] + \nu = \sigma^2(X) + \nu,$$

where ν is the projection error.
Then

$$\begin{aligned}
\mathbb{E}\left[e^2 - h(X)\right] &= \mathbb{E}\left[\nu + \sigma^2(X) - h(X)\right]^2 \\
&= \mathbb{E}\left[\nu^2\right] + \mathbb{E}\left[\sigma^2(X) - h(X)\right]^2 \\
&\geqslant \mathbb{E}[\nu^2]
\end{aligned}$$

where the equality can be attained when $h(X)$ is chosen to be $\sigma^2(X)$.

□

PROBLEM 2.16

PROOF. Following Theorem 2.9 of Hansen's book, we have coefficients of the best linear predictor:

$$\begin{aligned}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} &= \left\{\mathbb{E}\left[\begin{pmatrix}1\\ X\end{pmatrix}\begin{pmatrix}1 & X\end{pmatrix}\right]\right\}^{-1}\mathbb{E}\left[\begin{pmatrix}1\\ X\end{pmatrix}y\right]\\ &= \begin{bmatrix}1 & \mathbb{E}(X)\\ \mathbb{E}(X) & \mathbb{E}(X^2)\end{bmatrix}^{-1}\begin{bmatrix}\mathbb{E}(Y)\\ \mathbb{E}(XY)\end{bmatrix} \end{aligned} \tag{1}$$

Since $f(x, y)$ is given, it is not hard to compute $\mathbb{E}(X) = \mathbb{E}(Y) = 5/8$, $\mathbb{E}(X^2) = 7/15$ and $\mathbb{E}(XY) = 3/8$. Plugging those values and computing the inverse of the matrix in Eq. (1), we obtain coefficients:

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}55/73\\ -15/73\end{pmatrix},$$

and the best linear predictor

$$\mathscr{P}(Y \mid X = x) = \left(\frac{55}{73} - \frac{15}{73}x\right)\mathbb{1}_{[0,1]}(x), \tag{2}$$

where $\mathbb{1}_A(\cdot)$ denotes the indicator function of x for some set A such that $\mathbb{1}_A(x) = 1$ if $x \in A$.

The conditional mean function $m(x)$ is, by definition

$$m(x) \equiv \mathbb{E}(Y \mid X = x) = \int_0^1 y f(y \mid x)\,\mathrm{d}y = \int_0^1 y \cdot \frac{f(x, y)}{f(x)}\,\mathrm{d}y,$$

Note that $f(y \mid x), f(x, y)$ and $f(x)$ are shorthand for the corresponding conditional, joint and marginal densities involving X and Y, evaluated at $X = x$ and $Y = y$.

The marginal density $f(x)$ then can be obtained by

$$f(x) = \int_0^1 f(x, y)\,\mathrm{d}y = \frac{1}{2}\left(3x^2 + 1\right)\mathbb{1}_{[0,1]}(x)$$

Then it is not hard to have

$$m(x) = \left(\frac{6x^2 + 3}{12x^2 + 4}\right)\mathbb{1}_{[0,1]}(x)\ . \tag{3}$$

By Eq. (2) and Eq. (3), it is obvious that $\mathscr{P}(Y \mid X = x)$ and $m(x)$ do not coincide.

Their discrepancy arises from the fact that the best linear predictor $\mathscr{P}(Y \mid X = x)$, by construction, places a linear restriction on the relationship between Y and X, implying only their first and second moments matter, whereas the conditional mean function $m(x)$ does not rely on any specific parametric form and thus makes full use of information on the joint distribution of x and y. $\square$

PROBLEM 2.19

Proof. By definition, the loss function

$$\begin{aligned} \mathrm{d}(\boldsymbol{\beta}) &= \mathbb{E}\left[(m(\boldsymbol{X}) - \boldsymbol{X}'\boldsymbol{\beta})^2\right] \\ &= \mathbb{E}\left[m^2(\boldsymbol{X})\right] + \boldsymbol{\beta}'\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}'\right]\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbb{E}\left[\boldsymbol{X}m(\boldsymbol{X})\right]. \end{aligned}$$

FOC w.r.t. $\boldsymbol{\beta}$ gives

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{d}(\boldsymbol{\beta}) = 2\mathbb{E}[\boldsymbol{X}\boldsymbol{X}']\boldsymbol{\beta} - 2\mathbb{E}[\boldsymbol{X}m(\boldsymbol{X})].$$

Therefore,

$$\boldsymbol{\beta} = \left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}']\right)^{-1}\mathbb{E}\left[\boldsymbol{X}m(\boldsymbol{X})\right]. \tag{4}$$

By the *law of iterated expectations*, we know

$$\begin{aligned} \mathbb{E}[\boldsymbol{X}Y] &= \mathbb{E}[\mathbb{E}[\boldsymbol{X}Y \mid \boldsymbol{X}]] \\ &= \mathbb{E}[\boldsymbol{X}\underbrace{\mathbb{E}[Y \mid \boldsymbol{X}]}_{\equiv m(\boldsymbol{X})}]. \\ &= \mathbb{E}[\boldsymbol{X}m(\boldsymbol{X})]. \end{aligned}$$

Plugging the above equality into Eq. (4) yields

$$\boldsymbol{\beta} = \left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}']\right)^{-1}\mathbb{E}\left[\boldsymbol{X}Y\right]. \tag{5}$$

So, expressions Eq. (4) and Eq. (5) are equivalent. □