Question 1. Decompose the Expected Prediction Error into three parts: Irreducible error, squared Bias, and Variance

Solution: We treat $\boldsymbol{x}_0$ and f as fixed, i.e. non-random. Note that Y_0 is not in the original sample, and so $\widehat{f}$ is independent of both Y_0 and ε_0. We have:

$$\begin{aligned}\text{EPE} &= E\left[\left(Y_0 - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(f(\boldsymbol{x}_0) + \varepsilon_0 - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(\varepsilon_0 + \left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right)^2\right] \\ &= E\left[\varepsilon_0^2\right] + 2E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] + E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right].\end{aligned}$$

Note that $2E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] = 0$, because ε_0 is independent from $\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)$, and $E\left[\varepsilon_0\right] = 0$, which implies $E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] = E\left[\varepsilon_0\right]E\left[f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right] = 0$. Also note that, by definition, $\text{Var}\left(\varepsilon_0\right) = E\left[\varepsilon_0^2\right] - \left(E\left[\varepsilon_0\right]\right)^2$. Thus, $E\left[\varepsilon_0^2\right] = \text{Var}\left(\varepsilon_0\right) = \sigma^2$. Consequently,

$$\text{EPE} = \sigma^2 + E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = \text{Irreducible error } + \text{ Reducible error}.$$

Now we focus on the Reducible error:

$$\begin{aligned}&E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right] + E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] \\ &= \left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2 + 2E\left[\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)\right] \\ &+ E\left[\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right].\end{aligned}$$

The middle term is zero again. To see this note that the only random component in this term is $\widehat{f}(\boldsymbol{x}_0)$, and $E\left[E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right] = 0$. Hence, the Reducible error is:

$$\begin{aligned}&\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2 + E\left[\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] \\ &= \left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - f(\boldsymbol{x}_0)\right)^2 + E\left[\left(\widehat{f}(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2\right] \\ &= \text{Bias}^2\left(\widehat{f}(\boldsymbol{x}_0)\right) + \text{Var}\left(\widehat{f}(\boldsymbol{x}_0)\right)\end{aligned}$$

Putting it all together,

$$\text{EPE} = \sigma^2 + \text{Bias}^2\left(\widehat{f}(\boldsymbol{x}_0)\right) + \text{Var}\left(\widehat{f}(\boldsymbol{x}_0)\right).$$

Question 2. Show that the OLS estimator is unbiased, i.e., derive the following:

$$E\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}.$$

Treat the x values as fixed (i.e. non-random) and use the formula for the OLS estimator.

Solution: Recall that the expected value of the error terms in the MLR model is zero. We will make use of the following formulas (and all the corresponding notation) from Lecture 3:

$$\begin{aligned} &\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \qquad \text{and} \\ &\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \end{aligned}$$

In the following expected value calculations, non-random matrixes are treated as constants, which we can be factored out of the expected values. We have:

$$\begin{aligned} E\widehat{\boldsymbol{\beta}} &= E\Big[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}\Big] \\ &= E\Big[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\Big] \\ &= E\Big[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}\Big] + E\Big[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}\Big] \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta}. \end{aligned}$$

Question 3. Let $y_1, \ldots, y_n$ be a sample from a distribution with the density function $p(y;\theta) = \theta y^{\theta-1}$ for $0 < y < 1$, where $\theta > 0$.

Find $\widehat{\theta}$, the maximum likelihood estimator of θ.

Compute $\widehat{\theta}$ for the sample $y_1 = 0.35$, $y_2 = 0.28$, $y_3 = 0.91$.

Solution: The likelihood function is

$$\begin{aligned}\ell(\theta) &= p(y_1;\theta)p(y_2;\theta)\ldots p(y_n;\theta) \\ &= \prod_{i=1}^{n} \theta y_i^{\theta-1} \\ &= \theta^n \prod_{i=1}^{n} y_i^{\theta-1}.\end{aligned}$$

Taking the natural log:

$$L(\theta) = \log(\ell(\theta)) = n\log(\theta) + (\theta - 1)\sum_{i=1}^{n} \log(y_i).$$

The first derivative is

$$\frac{dL(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(y_i).$$

The first derivative is zero at $\widehat{\theta}$:

$$\frac{n}{\widehat{\theta}} + \sum_{i=1}^{n} \log(y_i) = 0.$$

Thus,

$$\widehat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log(y_i)}.$$

For the sample $0.35, 0.28, 0.91$, we have:

$$\widehat{\theta} = \frac{-3}{\log(0.35) + \log(0.28) + \log(0.91)} = 1.24.$$