EML Assignment 1 Problem 2

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

Question 1

Definitions

$$
\begin{aligned}
f(x_0) &\equiv \text{is the value of the true distribution function at } x_0 \\
\hat{f}(x) &\equiv \text{is the random variable of the value of an approximation of } f(x) \text{ at } x_0 \\
\mathrm{Var}(\hat{f}(x_0)) &= \mathbb{E}\left[(\hat{f}(x_0) - \mathbb{E}[\hat{f}(x_0)])^2\right] \\
\mathrm{Bias}(\hat{f}(x_0)) &= \mathbb{E}(\hat{f}(x_0) - f(x_0))
\end{aligned}
$$

Assumptions

$$y = f(x) + \epsilon$$

Where the random noise term *epsilon* is independent and has a zero expected value.

$$
\begin{aligned}
\mathbb{E}(\epsilon) &= 0 \\
\mathrm{Var}(\epsilon) &= \mathbb{E}(\epsilon^2)
\end{aligned}
$$

Proof

Let $\hat{y}_0 = \hat{f}(x_0)$.

$$
\begin{aligned}
\mathbb{E}\left[(y_0 - \hat{(}f(x_0))^2\right] &= \mathbb{E}\left[(y_0 - \hat{y}_0)^2\right] \\
&= \mathbb{E}\left[((f(x_0) - \hat{y}_0) + \epsilon)^2\right] && y_0 = f(x_0) + \epsilon \\
&= \mathbb{E}\left[(f(x_0) - \hat{y}_0)^2 + 2(f(x_0) - \hat{y}_0)\epsilon + \epsilon^2\right] \\
&= \mathbb{E}[(f(x_0) - \hat{y}_0)^2] + \mathbb{E}[(f(x_0) - \hat{y}_0)] \cdot \mathbb{E}(\epsilon) + \mathbb{E}(\epsilon^2) \\
&= \mathbb{E}[(f(x_0) - \hat{y}_0)^2] + \mathrm{Var}(\epsilon) && \mathbb{E}[\epsilon] = 0 \text{ and } \mathbb{E}[\epsilon^2] = \mathrm{Var}(\epsilon)
\end{aligned}
$$

Now, we focus on the first term.

$$
\begin{aligned}
\mathbb{E}\left[(f(x_0)-\hat{y}_0)^2\right] &= \mathbb{E}\left[((f(x_0)-\mathbb{E}[\hat{y}_0])+(\mathbb{E}[\hat{y}_0]-\hat{y}_0))^2\right] \\
&= \mathbb{E}\left[(f(x_0)-\mathbb{E}[\hat{y}_0])^2+2(f(x_0)-\mathbb{E}[\hat{y}_0])(\mathbb{E}[\hat{y}_0]-\hat{y}_0)+(\mathbb{E}[\hat{y}_0]-\hat{y}_0)^2\right] \\
&= (f(x_0)-\mathbb{E}[\hat{y}_0])^2+2(f(x_0)-\mathbb{E}[\hat{y}_0])(\mathbb{E}[\hat{y}_0]-\mathbb{E}[\hat{y}_0])+\mathbb{E}\left[(\mathbb{E}[\hat{y}_0]-\hat{y}_0)^2\right] \\
&= (f(x_0)-\mathbb{E}[\hat{y}_0])^2+\mathbb{E}\left[(\mathbb{E}[\hat{y}_0]-\hat{y}_0)^2\right] \\
&= \mathbb{E}\left[(\hat{y}_0-\mathbb{E}[\hat{y}_0])^2\right]+(\mathbb{E}[\hat{y}_0]-f(x_0))^2 \\
&= \mathbb{E}\left[(\hat{f}(x_0)-\mathbb{E}(\hat{f}(x_0))^2\right]+\left[\mathbb{E}(\hat{f}(x_0)-f(x_0))\right]^2
\end{aligned}
$$

Where we used the fact that the term $(f(x_0)-\mathbb{E}[\hat{y}_0])$ is constant. Now, we substitute this result and use the earlier definitions for bias and variance:

$$
\begin{aligned}
\mathbb{E}[(y_0-\hat{f}(x_0)] &= \mathbb{E}\left[(\hat{f}(x_0)-\mathbb{E}(\hat{f}(x_0))^2\right]+\mathrm{Var}(\epsilon) \\
&= \mathrm{Var}(\hat{f}(x_0))+\mathrm{Bias}(\hat{f}(x_0))^2+\mathrm{Var}(\epsilon)
\end{aligned}
$$

Question 2

the reducible error is the controllable error, which can be reduced by modifying the parameters of the model you are able to reduce its value. However, on the other hand, the irreducible error is the error due to the noise nature, which cannot be reduced regardless of how you change the parameters of the model. The model still won't be able to generalize, since the noise is a natural property of the data sampling.