

Elements of Machine Learning

Assignment 1 - Problem 2

Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

November 11, 2021

2 Bias-variance trade-off

Q. Prove that the expected test mean square error (MSE), for a given value x_0 , can always be decomposed into the sum of three fundamental quantities: the variance of $\hat{f}(x_0)$, the squared bias of $\hat{f}(x_0)$ and the variance of the error terms ϵ . Please note that you should prove both qualities (both equals relation), enlist all assumptions and define all parameters and variables.

Solution

We have

- y_0 : Given n training data points $i = 1, 2, \dots, n$, it holds $y = f(x) + \epsilon$, where ϵ is a noise or error term.
- $f(x_0)$: true function
- $\hat{f}(x_0)$: estimated function

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

Since f is independent $\mathbb{E}[f(x_0)] = f$ and $\mathbb{E}[(f(x_0) - \mathbb{E}[\hat{f}(x_0)])] = (f(x_0) - \mathbb{E}[\hat{f}(x_0)])$ since the quantity inside is difference between constant ($f(x_0)$) (because it is a given data) and average prediction ($\mathbb{E}[\hat{f}(x_0)]$). Also, ϵ is error thus $\mathbb{E}[\epsilon] = 0$

$$\begin{aligned}
\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2] \\
&\quad + \underbrace{2(f(x_0) - \mathbb{E}[\hat{f}(x_0)])\mathbb{E}[\epsilon]}_{0(\text{As } \mathbb{E}[\epsilon]=0)} \\
&\quad + \underbrace{2\mathbb{E}[\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0)]\mathbb{E}[\epsilon]}_{0(\text{As } \mathbb{E}[\epsilon]=0)} \\
&\quad + 2\mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))(f(x_0) - \mathbb{E}[\hat{f}(x_0)])] \\
\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2] \\
&\quad + 2\mathbb{E}[\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0)](f(x_0) - \mathbb{E}[\hat{f}(x_0)]) \\
\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2] \\
&\quad + 2 \left(\mathbb{E}[\mathbb{E}[\hat{f}(x_0)]] - \mathbb{E}[\hat{f}(x_0)] \right) (f(x_0) - \mathbb{E}[\hat{f}(x_0)]) \\
\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2] \\
&\quad + \underbrace{2(\mathbb{E}[\hat{f}(x_0)] - \mathbb{E}[\hat{f}(x_0)])(f(x_0) - \mathbb{E}[\hat{f}(x_0)])}_0 \\
\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2]
\end{aligned}$$

Now we know that,

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

Using these, we have

$$\mathbb{E} \left[(y_0 - \hat{f}(x_0)) \right]^2 = (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + \mathbb{E}[\epsilon^2] + \mathbb{E}[(\mathbb{E}[\hat{f}(x_0)] - \hat{f}(x_0))^2] \quad (1)$$

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

Q. Explain in your own words irreducible and reducible error as well as their difference.

Solution

Irreducible error is the noise in the function that describes the given data points. We call it irreducible because it is independent of any quantity and thus cannot be minimized. While, the reducible error is $E[f - \hat{f}]^2$. It is a squared error between the true value and the estimated value of f . We call it reducible error because the estimated function \hat{f} can be reformed to minimize the gap between true and estimated values.