FYS-STK4155
week37 exercises

yevheniv

September 2023

1 Bias-varinace tradeoff

Here are some definitions, that I will be using while derivation

$$y = f(x) + \varepsilon. \tag{1}$$

$$\text{MSE} = \mathbb{E}[(y - \tilde{y})^2]. \tag{2}$$

$$\text{BIAS} = \mathbb{E}[(y - \mathbb{E}[\tilde{y}])^2]. \tag{3}$$

Now, let's derive the relation between bias and variance:

$$\begin{aligned}
MSE &= \mathbb{E}[(y - \tilde{y})^2] \\
&= \mathbb{E}[(f + \varepsilon - \tilde{y}^2] \\
&= \mathbb{E}[(f + \varepsilon - \tilde{y} + \mathbb{E}[\tilde{y}] - \mathbb{E}[\tilde{y}])^2] \\
&= \mathbb{E}[(f - \mathbb{E}[\tilde{y}])^2] + \mathbb{E}[\varepsilon^2] + \mathbb{E}[(\mathbb{E}[\tilde{y}] - \tilde{y})^2] + 2\mathbb{E}[(f - \mathbb{E}[\tilde{y}])\varepsilon] \\
&\quad + \mathbb{E}[(\varepsilon\mathbb{E}[\tilde{y}] - \tilde{y})] + 2\mathbb{E}[(\mathbb{E}[\tilde{y}] - \tilde{y})(f - \mathbb{E}[\tilde{y}])] \\
&= (f - \mathbb{E}[\tilde{y}])^2 + \mathbb{E}[\varepsilon^2] + \mathbb{E}[(\mathbb{E}[\tilde{y}] - \tilde{y})^2] + 2(f - \mathbb{E}[\tilde{y}])\mathbb{E}[\varepsilon] \\
&\quad + 2\mathbb{E}[\varepsilon]\mathbb{E}[\mathbb{E}[\tilde{y} - \tilde{y}] + 2\mathbb{E}[\mathbb{E}[\tilde{y}] - \tilde{y}](f - \mathbb{E}[\tilde{y}]) \\
&= (f - \mathbb{E}[\tilde{y}])^2 + \mathbb{E}[\varepsilon^2] + \mathbb{E}[(\mathbb{E}[\tilde{y}] - \tilde{y})^2] \\
&= (f - \mathbb{E}[\tilde{y}])^2 + \text{var}[\varepsilon] + \text{var}[\tilde{y}] \\
&= (\text{Bias}[\hat{f}])^2 + \text{var}[\varepsilon] + \text{var}[\tilde{y}] \\
&= (\text{Bias}[\tilde{y}])^2 + \sigma^2 + \text{var}[\tilde{y}].
\end{aligned}$$

So, the relation is

$$(\text{Bias}[\tilde{y}])^2 + \sigma^2 + \text{var}[\tilde{y}]. \tag{4}$$

From this equation we can interpret the meaning of the bias and variance terms. Bias term represents "how far spread" prediction points and actual points. Variance term represents "how far spread" prediction points. Variance error is irreducible, while bias is reducible. So, for ideal model bias and variance will be very small.