1 Deriving the bias-variance trade-off equation

Let us write the true relationship between Y and x as

$$Y(x) = f(x) + \epsilon, \tag{1}$$

where ϵ has a zero mean, $E[\epsilon] = 0$, and a variance $E[\epsilon^2] = \sigma_\epsilon^2$.

We now fit a function to data from this underlying relation and get an estimating function, $\hat{f}(x)$. We are now interested in what the error is in our prediction at some point x. To make progress here we first need to specify what we mean by error — in other words we need to specify a loss, or error, function. Let us here use the squared error, so that we can write

$$\text{Error}(x) = \left(Y(x) - \hat{f}(x)\right)^2. \tag{2}$$

However this is a random quantity (since $\hat{f}$ is), so to make progress we need to take the expectation value of this

$$\text{Err}(x) = E\left[\text{Error}(x)\right] = E\left[(Y(x) - \hat{f}(x))^2\right]. \tag{3}$$

This is then the error that we want to investigate further. Before we continue it is convenient to recall how we define the bias and the variance of the estimator:

$$\text{Bias}\left(\hat{f}(x)\right) = E\left[\hat{f}(x)\right] - f(x) \tag{4}$$

and

$$\text{Var}\left(\hat{f}(x)\right) = E\left[\left(f(x) - E\left[\hat{f}(x)\right]\right)^2\right]. \tag{5}$$

In what follows I will suppress the argument x for simplicity, it should be obvious where it has to be inserted.

We can return to equation (3). If we expand this, we get

$$E\left[(Y - \hat{f})^2\right] = E\left[(f - \hat{f})^2\right] + 2E\left[(f - \hat{f})\epsilon\right] + E\left[\epsilon^2\right]. \tag{6}$$

The last term on the right is easy since we had defined $E[\epsilon^2] = \sigma_\epsilon^2$. The second term on the right is also easily. Because for uncorrelated random variables X and Y, $E[XY] = E[X]E[Y]$, and since $E[\epsilon] = 0$, the second term is zero.

This then leaves us only with $E[(f - \hat{f})^2]$. To make progress here, we add and subtract $E[\hat{f}]$ inside the parenthesis (because this term is needed for the bias and variance definitions):

$$E\left[(f - \hat{f})^2\right] = E\left[(f - E[\hat{f}] + E[\hat{f}] - \hat{f})^2\right] \tag{7}$$

$$= E\left[(f - E[\hat{f}])^2\right] + 2E\left[\left(f - E[\hat{f}]\right)\left(E[\hat{f}] - \hat{f}\right)\right] + \tag{8}$$

$$+ E\left[(\hat{f} - E[\hat{f}])^2\right]. \tag{9}$$

Here we recognised the last term on the right as $\mathrm{Var}(\hat{f})$ from equation (5), and the first term is $\mathrm{Bias}(\hat{f})^2$ from equation (4) and the realisation that the argument to the expectation value is not a random variable so the expectation operation just returns the argument.

That leaves only the middle term on the right, but we can see that this is of the form $E[aX]$ with a a constant and X a random variable. In this case $X = E[\hat{f}] - \hat{f}$ and by definition of $E[\hat{f}]$ we have $E[X] = 0$. Thus the middle term disappear.

That leaves us with

$$\mathrm{Err}(x) = \sigma_\epsilon^2 + \mathrm{Bias}^2(x) + \mathrm{Var}(x), \tag{10}$$

which is the bias-variance trade-off equation.