Bias-Variance Decomposition

Definitions

Bias$(\hat{y}) = E[\hat{y} - y] = E[\hat{y}] - E[y]$
$Var(\hat{y}) = E[(\hat{y} - E[\hat{y}])^2]$
$Var(y) = E[(y - E[y])^2] = \sigma^2$ (noise or irreducibly uncertainty)

Draw many training sets. On each one compute the average difference between average prediction and average truth; this gives the bias.
Note that for linear regression $E[y] = xw$ and that hidden inside $\hat{y}$ is the model complexity and the training set size.

Error (Sum of squared error)
$E[(y - \hat{y})^2] = Bias(\hat{y})^2 + Var(\hat{y}) + \sigma^2$

Proof sketch

$$
\begin{aligned}
E[(y - \hat{y})^2] &= E[(y - E[\hat{y}] + E[\hat{y}] - \hat{y})^2] \\
&= E[((y - E[\hat{y}])^2 + (E[\hat{y}] - \hat{y})^2 + 2(y - E[\hat{y}])(E[\hat{y}] - \hat{y})] \\
&= \text{ bias+noise } + \text{ variance } + \text{ this term vanishes}
\end{aligned}
$$

Now do a similar trick to pull out the bias and noise from within the first term.

$$
\begin{aligned}
E[(y - E[\hat{y}])^2] &= E[(y - E[y] + E[y] - E[\hat{y}])^2] \\
&= E[(y - E[y])^2] + E[(y - \hat{y})]^2 + \text{cross term that vanishes} \\
&= \text{irreducible uncertainty } + \text{ bias}^2
\end{aligned}
$$