Bias - Variance Tradeoff

True data generated from noisy model $\vec{y} = f(x) + \vec{\varepsilon}$ with $\vec{\varepsilon} \approx N(0, \sigma^2)$

Approximate f by the model: $\tilde{y} = X\hat{\beta}$

with $\hat{\beta}$ optimal parameters minimizing $C(X, \beta) = \mathbb{E}[(y - \tilde{y})^2] = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2$

Therefore:

$$C = \mathbb{E}[(y - \tilde{y})^2]$$
$$= \mathbb{E}[(f + \varepsilon - \tilde{y})^2]$$
$$= \mathbb{E}[(f - \tilde{y})^2 + 2\varepsilon \cdot (f - \tilde{y}) + \varepsilon^2]$$
$$= \mathbb{E}[(f - \tilde{y})^2] + \mathbb{E}[\varepsilon^2]$$
$$= \mathbb{E}[(f - \tilde{y})^2] + \sigma^2$$

$\mathbb{E}[\varepsilon] = 0$

$var[\varepsilon] = \mathbb{E}[\varepsilon^2] - \mathbb{E}[\varepsilon]^2 = \sigma^2$

$\sigma^2 = \mathbb{E}[\varepsilon^2]$

Adding and substracting

$$= \mathbb{E}\left[\left(f + \mathbb{E}[\tilde{y}] - \tilde{y} - \mathbb{E}[\tilde{y}]\right)^2\right] + \sigma^2$$
$$= \mathbb{E}\left[\left((f - \mathbb{E}[\tilde{y}]) - (\tilde{y} - \mathbb{E}[\tilde{y}])\right)^2\right] + \sigma^2$$

Expanding

$$= \mathbb{E}\left[(f - \mathbb{E}[\tilde{y}])^2 + (\tilde{y} - \mathbb{E}[\tilde{y}])^2 - 2(f - \mathbb{E}[\tilde{y}]) \cdot (\tilde{y} - \mathbb{E}[\tilde{y}])\right] + \sigma^2$$
$$= \mathbb{E}[(f - \mathbb{E}[\tilde{y}])^2] + \mathbb{E}[(\tilde{y} - \mathbb{E}[\tilde{y}])^2] - 2\mathbb{E}[\underbrace{(f - \mathbb{E}[\tilde{y}])}_{\text{NON-STOCHASTIC}} \cdot (\tilde{y} - \mathbb{E}[\tilde{y}])] + \sigma^2$$

$$C = \mathbb{E}[(f - \mathbb{E}[\tilde{y}])^2] + \mathbb{E}[(\tilde{y} - \mathbb{E}[\tilde{y}])^2] - 2(f - \mathbb{E}[\tilde{y}]) \cdot \underbrace{\mathbb{E}[\tilde{y} - \mathbb{E}[\tilde{y}]]}_{=0} + \sigma^2$$

Note $\mathbb{E}[y] = \mathbb{E}[f + \varepsilon] = f$

Then

$$C(X, \beta) = \mathbb{E}\left[(\mathbb{E}[y] - \mathbb{E}[\tilde{y}])^2\right] + \mathbb{E}\left[(\tilde{y} - \mathbb{E}[\tilde{y}])^2\right] + \sigma^2$$
$$= Bias[\tilde{y}] + var[\tilde{y}] + \sigma^2$$

σ^2: Unavoidable error arising from the error in the experimental data $\vec{y}$ (noisy model)

$Bias[\tilde{y}]$: Indicates how far the mean values of the experimental data $\vec{y}$ and the model $\tilde{y}$ are, on average.

$var[\tilde{y}]$: Measures the average distance that the values of the $\tilde{y}(x)$ model deviate from their mean value. Indicates how spread out the model's data is.