Quiz Problem 3.

Consider a model where the true relationship between a response variable Y and a predictor variable X is given by:

$$Y = f(X) + \epsilon$$

where f is the true underlying function, and ϵ is the random error term with $\mathbb{E}[\epsilon] = 0$, $\text{Var}(\epsilon) = \sigma^2$, and $\epsilon \perp X$.

Let $\{(X_n, Y_n)\}_{n=1}^N$ be training data generated (independently) from this model and then used to construct a predictor $\hat{f}$. Consider a fixed point x and let $y = f(x) + \epsilon$ be a point generated according to our model.

For this fixed point x, define the MSE to be

$$\mathbb{E}[(y - \hat{f}(x))^2]$$

Derive an expression for the MSE in terms of bias and variance using the following definitions:

- Bias:
$$\text{Bias}(\hat{f}(x)) = \mathbb{E}[\hat{f}(x)] - f(x)$$
- Variance:
$$\text{Var}(\hat{f}(x)) = \mathbb{E}[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2]$$

Hint: Argue that

1. $\mathbb{E}[(f(x) - \hat{f}(x))\epsilon] = 0$.
2. $\mathbb{E}[(f(x) - \mathbb{E}[\hat{f}(x)])(\mathbb{E}[\hat{f}(x)] - \hat{f}(x))] = 0$