$$R(f) = \mathbb{E}(Y - f(X))^2 = \mathbb{E}(Y - \eta(X) + \eta(X) - f(X))^2$$

$$= \underbrace{\mathbb{E}(Y - \eta(X))^2}_{R^*} + \mathbb{E}(\eta(X) - f(X))^2 + 2\,\underbrace{\mathbb{E}(Y - \eta(X))(\eta(X) - f(X))}$$

$$= \mathbb{E}\Big[\mathbb{E}\big[(Y - \eta(X))(\eta(X) - f(X)) \mid X\big]\Big]$$

$$= \mathbb{E}\Big[\mathbb{E}[Y - \eta(X) \mid X]\,(\eta(X) - f(X))\Big]$$

$$= \mathbb{E}\Big[\underbrace{\mathbb{E}[Y - \mathbb{E}[Y \mid X] \mid X]}_{0}\,(\eta(X) - f(X))\Big] = 0$$

Suppose that f is random $(\hat{f})$ because it comes from training data.

$$\mathbb{E}(\hat{f}(X) - \eta(X))^2 = \mathbb{E}(\hat{f}(X) - \mathbb{E}[\hat{f}(X) \mid X])^2 + \mathbb{E}(\mathbb{E}[\hat{f}(X) \mid X] - \eta(X))^2$$

$$+ (\text{cross-term } \cancel{\text{that vanishes}} = 0)$$

$$= \mathbb{E}\big[\mathrm{Var}(\hat{f}(X) \mid X)\big] + \mathbb{E}\big[\mathrm{Bias}^2(\hat{f}(X) \mid X)\big]$$

w/ $\mathrm{Bias}(\hat{f}(X) \mid X) = \mathbb{E}[\hat{f}(X) \mid X] - \eta(X)$

<u>prop</u> $R(\hat{f}) = R^* + \mathbb{E}\big[\mathrm{Bias}^2(\hat{f}(X) \mid X) + \mathrm{Var}(\hat{f}(X) \mid X)\big]$