# Mean square error (MSE)

$\hat{f}(x)$: prediction function estimated with a set of data samples

$E_D[\hat{f}(x)]$: average of model due to different sample set

$$MSE = (f(x) - \hat{f}(x))^2$$

$$= (f(x) - E_D[\hat{f}(x)] + E_D[\hat{f}(x)] - \hat{f}(x))^2$$

$$= (f(x) - E_D(\hat{f}(x)))^2 + (E_D[\hat{f}(x)] - \hat{f}(x))^2$$

$$\underbrace{- 2(f(x) - E_D(\hat{f}(x)))(E_D[\hat{f}(x)] - \hat{f}(x))}_{=0} \; *$$

Applying $E_D$ to both sides.

$$E_D(f(x) - \hat{f}(x))^2 = \underbrace{(f(x) - E_D[\hat{f}(x)])^2}_{\text{bias}^2} + \underbrace{E_D[E_D[\hat{f}(x)] - \hat{f}(x))^2]}_{\text{Variance}}$$

the $E_D$ of this is equal to this since $f(x)$ is a constant

same reasoning

* this part of the equation is equal to zero because

$$E_D[2(f(x) - E_D[\hat{f}(x)])(E_D[\hat{f}(x)] - \hat{f}(x))]$$

$$= 2(f(x) - E_D[\hat{f}(x)])\, E_D[E_D[\hat{f}(x)] - \hat{f}(x)]$$

$$= 2(f(x) - E_D[\hat{f}(x)])(\underbrace{E_D[\hat{f}(x)] - E_D[\hat{f}(x)]}_{0})$$

$$= 0$$