Suppose we develope $\hat{f}$ from a training dataset $S = \{(\vec{X}_1, Y_1),$

$\vdots$

$(\vec{X}_n, Y_n)\}$

Y = course grade (%)

X_{Q1} = Quiz 1 grade

$$Y = 0.5X_m + 0.15X_Q + 0.15X_{HW} + 0.2X_F$$

X_{Q1}

$$\hat{Y} = \hat{f}(X_{Q1}) \qquad \therefore \qquad Y = f(X_{Q1}) + \varepsilon$$

random variable

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

assume $E_{\varepsilon}[\varepsilon] = 0$

then $Var(\varepsilon) = E_{\varepsilon}\varepsilon^2 - \cancel{[E(\varepsilon)]^2}^{\,0}$

$Y - \hat{Y}$

$Y - \hat{f}(X)$

$f(X) + \varepsilon - \hat{f}(X)$

all random variables

Y, X, ε, S

How good is $\hat{f}$?

Test it with (X, Y)

$$\hat{f}(X) = \hat{Y}$$

Is Y close to $\hat{Y}$? $\hat{Y} - Y$

$$E_{X,S,\varepsilon,Y}(Y-\hat{Y})^2$$

expected test squared prediction error

Bias-Variance decomposition

S is independent of X, Y

"training" " " " test

X is independent of ε

$$E_{X,S,\varepsilon,Y}\left[Y-\hat{Y}\right]^2$$

$$= E_{X,S,\varepsilon,Y}\left[f(X)+\varepsilon-\hat{f}_S(X)\right]^2$$

$$\overset{①}{=} E_{X,S,\varepsilon}\left[f(X)+\varepsilon-\hat{f}_S(X)\right]^2$$

$$\overset{②}{=} E_X E_S E_\varepsilon\left[f(X)+\varepsilon-\hat{f}_S(X)\right]^2$$

$$= E_X E_S E_\varepsilon\left(\left(f(X)-\hat{f}_S(X)\right)+(\varepsilon)\right)^2$$

FOIL

$$= E_X E_S E_\varepsilon\left[\left(f(X)-\hat{f}_S(X)\right)^2+2\varepsilon\left(f(X)-\hat{f}_S(X)\right)+\varepsilon^2\right]$$

$$\underbrace{E_X E_S\left(f(X)-\hat{f}_S(X)\right)^2}_{E_X\left[\left[f(X)-E_S\hat{f}_S(X)\right]^2\right]+Var_S\,\hat{f}_S(X)} + \underbrace{E_X E_S E_\varepsilon 2\varepsilon\left(f(X)-\hat{f}_S(X)\right)}_{\to 0} + E_X E_S \underbrace{E_\varepsilon\left[\varepsilon^2\right]}_{Var(\varepsilon)}$$

$$+Var_\varepsilon(\varepsilon)$$

$$= E_{Y,X,S,\varepsilon}\left[Y-\hat{Y}\right]^2$$

$$E_S\left[f(X)-\hat{f}_S(X)\right]^2$$

$$=E_S\left[f^2(X)-2f(X)\hat{f}_S(X)+\hat{f}_S^2(X)\right]$$

$$=E_S f^2(X)-2f(X)E_S\hat{f}(X)+E_S\left[\hat{f}_S^2(X)\right]$$

$$=f^2(X)-2f(X)E_S\left[\hat{f}_S(X)\right]+E_S\left[\hat{f}_S^2(X)\right]+\left[E_S\hat{f}_S(X)\right]^2-\left[E_S\hat{f}_S(X)\right]^2$$

$$\left[f(X)-E_S\hat{f}_S(X)\right]^2+\mathrm{Var}_S\left(\hat{f}_S(X)\right)$$

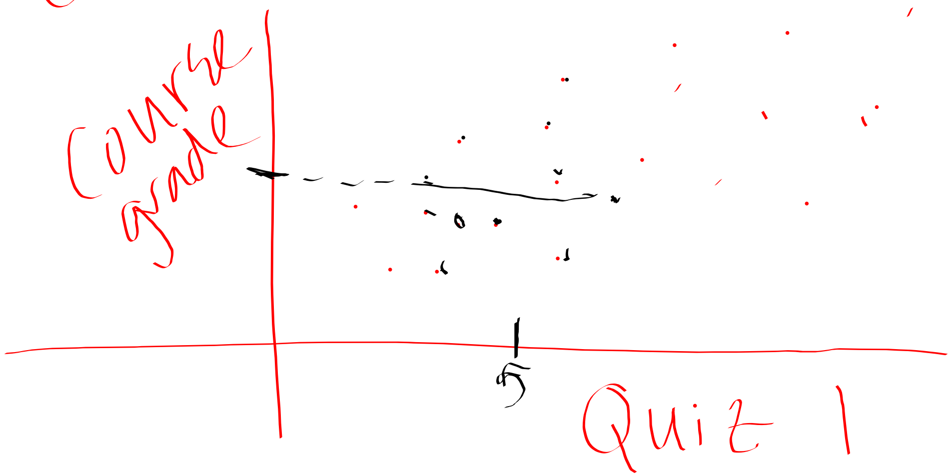
$Var_\varepsilon(\varepsilon)$ = irreducible error

$E_X Var_S \hat{f}_S(X)$ = spread of estimated f over different training sets, averaged over test points

$E_X[\text{bias}]^2$, bias $= f(X) - E_S \hat{f}_S(X)$

KNN for regression

$K = 10$



KNN

$$\hat{Y} = \sum_{i \in N_0} \frac{Y_i}{K}$$

Where N_0 is the set of K nearest points to X_0

KNN for regressio