Notation ~> See course book p. 9-12

p: # of predictors in our model:

Matrix

$$X_{n \times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & & x_{2p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

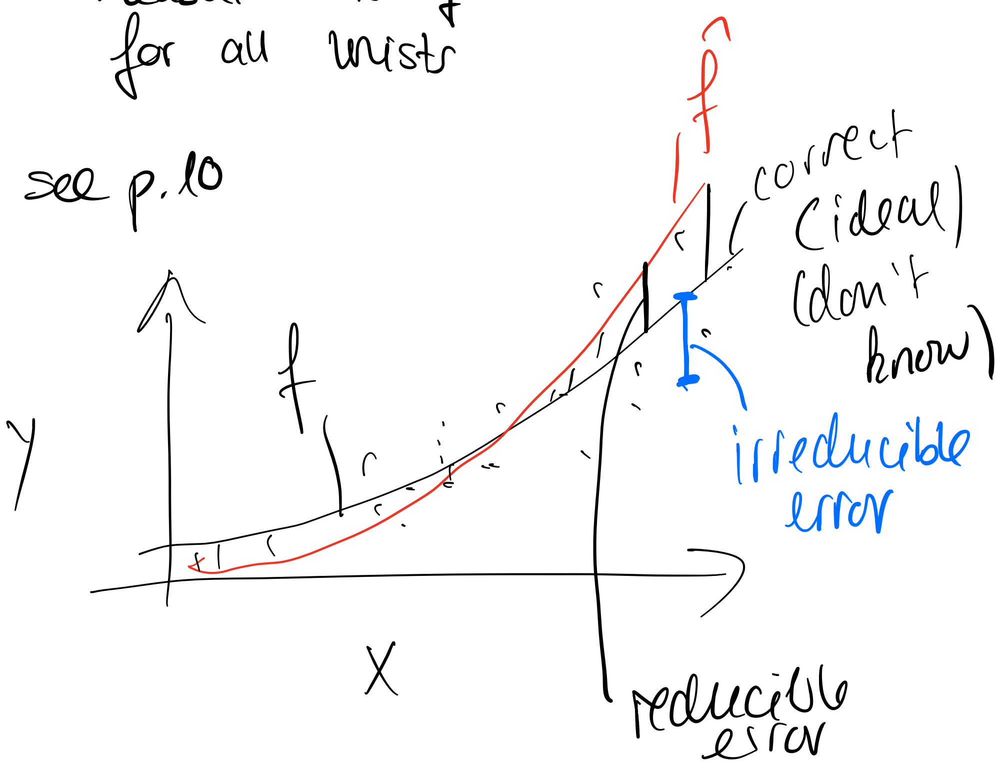
X_1^T all meas. for unit 1

n units

$\underbrace{x_{11}, \dots, x_{n1}}_{X_1}$ ... $\underbrace{x_{1p}, \dots, x_{np}}_{X_p}$

X_1 → measurements of variable 1 for all units

see p. 10



Reducible vs. irreducible error

Assume $\hat{f}$ and x are fixed

$$E\left[(Y-\underbrace{\hat{Y}}_{\hat{f}(x)})^2\right] = E\left[(f(X)+\varepsilon-\hat{f}(X))^2\right]$$

$$= E\left[(f(X)-\hat{f}(X))^2\right] + \overbrace{E\left[\varepsilon^2\right]}^{\text{var}(\varepsilon)}$$

$$\underbrace{+\,2\,E\left[\underbrace{(f(X)-\hat{f}(X))}_{\text{fixed}}\cdot\varepsilon\right]}_{=0\ \ (E(\varepsilon)=0)}$$

$$= (f(X)-\hat{f}(X))^2 + \text{var}(\varepsilon)$$

Bias-variance trade-off

Expected test MSE at x_0

$$E\left[(\underbrace{y_0}_{f(x_0)+\varepsilon}-\hat{f}(x_0))^2\right] = E\left[(f(x_0)+\varepsilon-\hat{f}(x_0))^2\right]$$

$$= \underbrace{E\left[(f(x_0))^2\right]} + \underbrace{E\left[\varepsilon^2\right]} + \overbrace{E\left[(\hat{f}(x_0))^2\right]}^{\text{var}(\hat{f}(x_0)) + E[\hat{f}(x_0)]^2}$$

$$-2E\left[f(x_0)\hat{f}(x_0)\right] + \underbrace{2E\left[\varepsilon\cdot f(x_0)\right]}_{\substack{=0\\ E(\varepsilon)=0}} - \underbrace{2E\left[\varepsilon\cdot \hat{f}(x_0)\right]}_{=0}$$

$$= f(x_0)^2 + Var(\varepsilon) + Var(\hat{f}(x_0)) + E\left[f(x_0)\right]^2$$

$$- 2E\left[f(x_0)\hat{f}(x_0)\right]$$

$$= \underbrace{Var(\varepsilon)} + \underbrace{Var(\hat{f}(x_0))} + \underbrace{\left(f(x_0) - E[\hat{f}(x_0)]\right)^2}$$