Notation ⇝ see course book 9-1

n (N) : number of units / observation

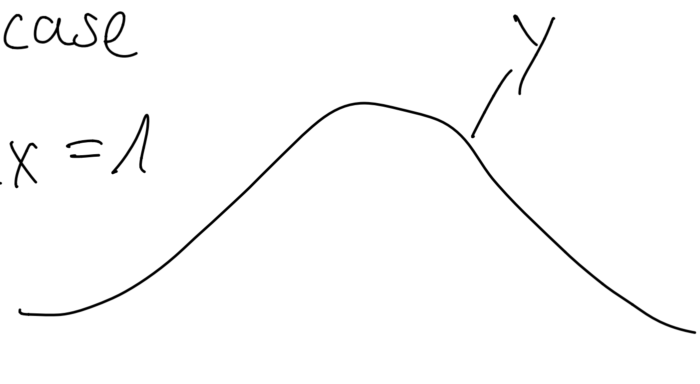
Y_i : response for unit i
random variable

continuous	binary	categorical
$N(\mu, \sigma^2)$	$Bin(1, p)$	multinomial

Distribution function f

Continuous case

$$\int f_Y(x)\,dx = 1$$



$$\mu_i = E[Y_i] = \int y_i f(y_i)\,dy_i \quad \text{(cont.)}$$

or $\sum_{y_i} y_i f(x_i)$ (discrete)

$$\sigma_i^2 = Var(Y_i) = E\left[(Y_i - E[Y_i])^2\right]$$
$$= E[Y_i^2] - E[Y_i]^2$$

Predictors

p: number of predictors in our model

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

X_1^T = meas. for unit 1

row vector

measurements of variable 1 (X_1) for all units

X_1

X_p

(see p. 10)

$$E[(Y-\hat{Y})^2] =$$

to be continued after the break

$$E[(f(X)+\varepsilon-\hat{Y})^2]$$

$\hat{f}(X)$

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

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

$$= \underbrace{(f(X)-\hat{f}(X))^2}_{\text{reducible}} + \overbrace{Var(\varepsilon)}^{\text{irred.}}$$

Bias-variance trade-off

Expected test MSE:

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

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

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

f fixed

= 0 (due to independence of ε and $E[\varepsilon] = 0$)

$$= f(x_0)^2 + Var(\varepsilon) + \overbrace{Var\left(\hat{f}(x_0)\right) + E\left(\hat{f}(x_0)\right)^2} - 2\,f(x_0)\,E\left(\hat{f}(x_0)\right)$$

$$= \underbrace{Var(\varepsilon)}_{\text{irreducible error}} + \underbrace{Var\left(\hat{f}(x_0)\right)}_{\text{variance}} + \underbrace{\left[f(x_0) - E\left(\hat{f}(x_0)\right)\right]^2}_{\text{squared bias}}$$