02/19/2018.

$y = \mathbb{R}$, p covariance

$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \}$ Linear assumption

$\mathcal{A}$: OLS

an element of the set.

$p = 1 \quad \mathcal{H} = \{ w_0 + w_1 x : w_0 w_1 \in \mathbb{R} \} \xrightarrow{\mathcal{A}: \text{minimize SSE of sqr error}} g(x) = b_0 + b_1 x$

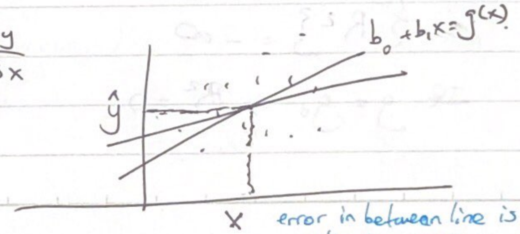
estimates of β_0 and β_1

the best assumption.

$h^*(x) = \beta_0 + \beta_1 x$

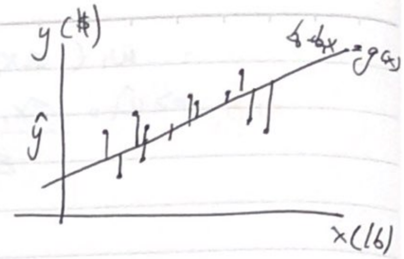
$b_0 = \bar{y} - r \dfrac{S_y}{S_x} \bar{x}$

$b_1 = r \dfrac{S_y}{S_x}$



error in between line is ostimation error.

Model performance metric how well does g predict?



$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y})^2$ (unit of y-symbol)

$MSE = \frac{1}{n-2} SSE$ (" " ")

mean sqd error

$RMSE = \sqrt{MSE}$

root

Approximate interpretation: g is **RMSE** off from y on average.

$g(x) \pm 2 \cdot RMSE$ is a 95% confidence set for y.

Empirical Rate (Rule of thumb)

null model →

Not parameter.

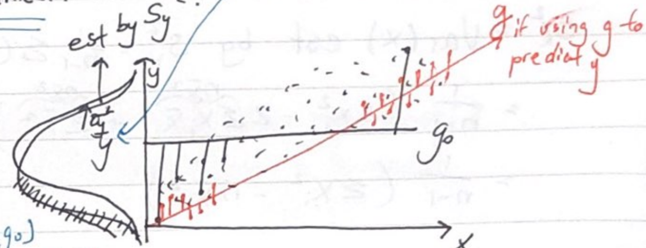
R^2: sample proportion of response variance explained.

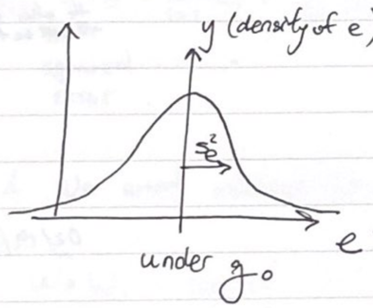
Consider the null model.

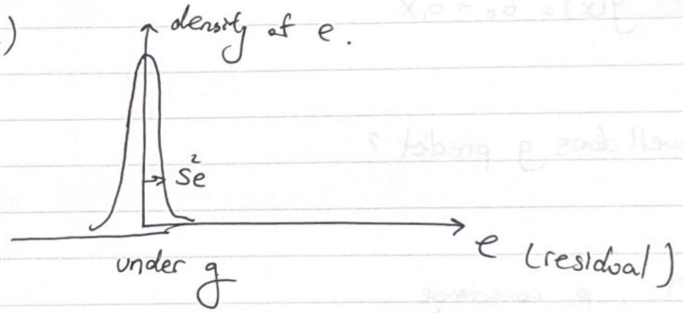
$g_0 = \bar{y}$ $\quad S_y^2 = \frac{1}{n-1} \underbrace{\sum (y_0 - \bar{y})^2}_{SSE_0}$

SSE_0 (sum of square error in g_0)

$= SST$ (sum of square total)







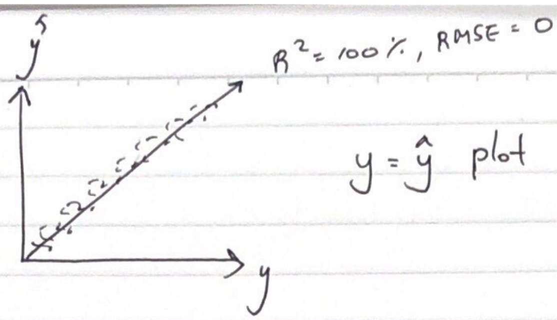
$$R^2 = \frac{\Delta s_e^2}{s_{e,0}^2} = \frac{s_{e,0}^2 - s_e^2}{s_{e,0}^2} = \frac{\frac{1}{n-1} \overset{SST}{\overset{\shortparallel}{SSE_0}} - \frac{1}{n-1} SSE}{\frac{1}{n-1} \underset{\underset{SST}{\shortparallel}}{SSE_0}} = \frac{\overbrace{SST - SSE}^{SSR}}{SST}$$

$$= \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

sup
max $\{R^2\} = 1$

inf
min $\{R^2\} = -\infty$ if you have a negative R^2 model need to check for mistake or very bad model.

If $g = g_0 \Rightarrow R^2 = 0$



$y = \hat{y}$ plot

ε = diff h^* and y

e = diff g and y. (residual.)

If R^2 goes up ↑ ⇒ RMSE goes ↓

If R^2 goes ~~up~~ down ↓ ⇒ RMSE goes ↑

Example:

y = # days of late payment.

$R^2 = 98\%$

RMSE = 25 days ⇒ empirical rule ± 50 days from y

Lab 4

$y \sim x$ → fit y into x.
"formula"

RMSE = σ (sigma).

Boston Housing data

phenomenon ⇒ price.

Linear model $p = 1$

one feature in y → $x_{rom} \in \{\text{red}, \text{green}\}$

$\qquad\qquad\qquad\downarrow\qquad\quad\downarrow$

coded $\qquad\quad 0 \qquad\quad 1$ ← indicator function $\mathbb{1}_{green}$

→ $x \in \{0, 1\}$ binary

$y \sim x$ $\mathcal{A}$: OLS $= b_0 + b_1 x$

⇑ proof

$$\hat{y} = g(x) = \begin{cases} \bar{y}_r & \text{if } x_{rom} = \text{red } (x = 0) \\ \bar{y}_g & \text{if } x_{rom} = \text{green. } (x = 1) \end{cases} = \bar{y}_r + (\bar{y}_g - \bar{y}_r)x$$