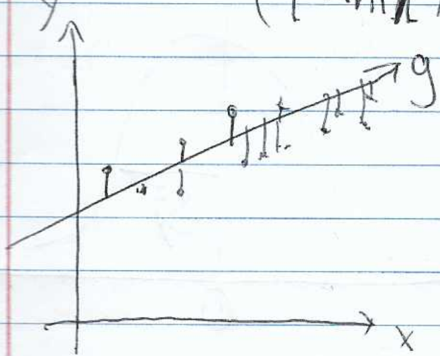
2/26/18

Lecture 8

$p=1$ $\quad \mathbb{D} = \left\{ \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\}$ $\quad \mathcal{Y} \subseteq \mathbb{R}$

$g \in \mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^2 \}$

$\vec{x} = [1, x]$



$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$

$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \text{argmin} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \in \mathbb{R}^2$ $\Big\}$ OLS "Ordinary least squares" where SSE is answer

$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - (w_0 + w_1 x_i))^2$

$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \not\Rightarrow \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \bar{y} - r\frac{s_y}{s_x}\bar{x} \\ r\frac{s_y}{s_x} \end{bmatrix}$ rare

$y_i = \underbrace{g(x_i)}_{\hat{y}_i} + e_i$

How well does this model do?

① $SAE := \sum |e_i|$

② $MAE := \frac{1}{n} \sum |e_i|$

$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

$s = \sqrt{s^2}$

③ $SSE = \sum e_i^2$

$MSE = \frac{1}{n-2} \sum e_i^2 = \frac{SSE}{n-2}$

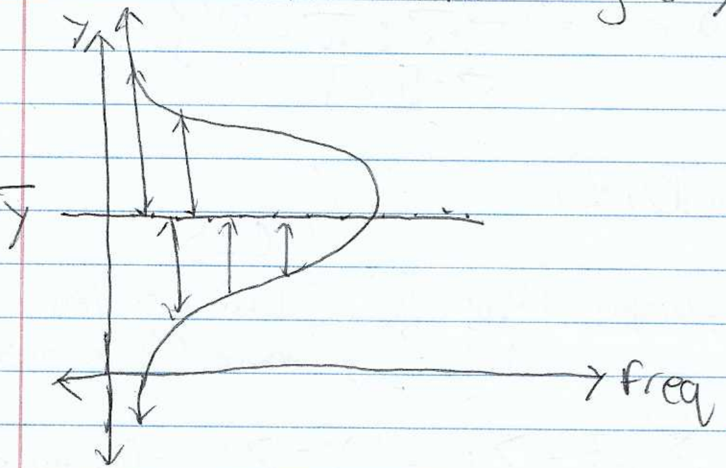
$RMSE = \sqrt{MSE}$ root mean sqd. error (most common metric that regression models report)

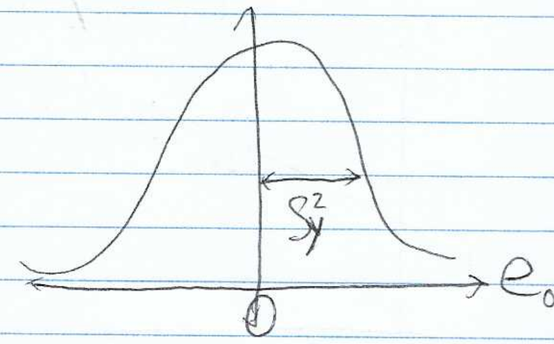
Confidence interval $CI_{M,95\%} := [\bar{x} \pm 2S_x]$ requires CLT (Central Limit Theorem)

Heuristic/empirical rule: $\pm 2RMSE \Rightarrow 95\%$ of predictive errors
rough rule of thumb

R^2: "Proportion of sample variance explained"

Consider the null model $g(x) = \bar{y}$ (If average price of used car is 10k You guess next used car is also 10k)

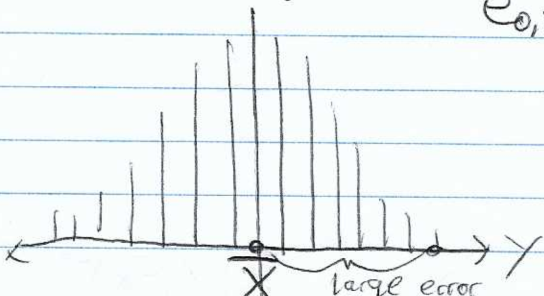




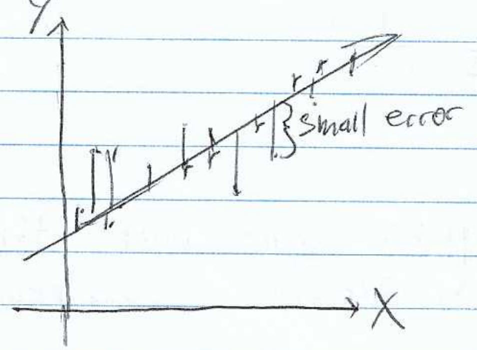
Using the null model you make mistakes

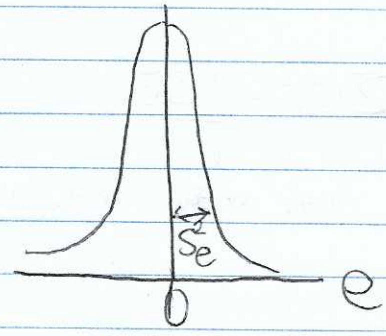
$$SST := SSE_0 := \sum \underbrace{(y_i - \bar{y})}_{e_{0,i}}{}^2 = (n-1)\, s_y^2$$

"sum of sqd. total"



after Modeling





$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum (e_i - \underbrace{\bar{e}}_{0})^2 = (n-1)\, s_e^2$$

$$R^2 = \frac{\overbrace{S_y^2 - S_e^2}^{\Delta S_y^2}}{S_y^2} = \frac{SSE_0 - SSE}{SSE_0} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = \text{"SSR"}$$

$\approx RSE = RMSE^2$

FYI

$$R^2 = \frac{S_y^2 - S_e^2}{S_y^2} \approx \frac{Var(Y) - Var(E)}{Var(Y)}$$

$R^2 \uparrow \Rightarrow RMSE \downarrow$

$R^2 \downarrow \Rightarrow RMSE \uparrow$

(Coefficient of determination FYI)

y: price of used car : $R^2 = 99\%$ unitless

but

RMSE = \$ 2000 units of y