Lec 8 Math 390.4 2/26/18

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

"Statistical learning" since there $\mathbb{D}$ and we want to create a g for prediction.

"regression" since $\mathcal{Y} \subset \mathbb{R}$ not $\{0, 1\}$ classification $\hat{y} = g(\vec{x}^*)$

$$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \} \overset{\text{for } p=1}{=} \{ w_0 + w_1 x_1 : w_0 \in \mathbb{R}, w_1 \in \mathbb{R} \}$$

$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$ $\mathcal{A}$ must "fit" w_0, w_1 two parameters

$\mathcal{A}$: "least squares regression" this requires solving the following problem:

$$\underset{w_0, w_1}{\text{argmin}} \; SSE = \sum_{i=1}^{n} \underbrace{\underbrace{\underbrace{(y_i - g(x_i))^2}_{(y_i - \hat{y}_i)^2}}_{(y_i - (w_0 + w_1 x_i))^2}}_{e_i^2}$$

We call these Least Squares estimates b_0, b_1

In many books, you will see them as $\hat{\beta}_0, \hat{\beta}_1$. I won't use this notation

$y_i = g(x_i) + e_i$ the amount left over

$$b_0 = \bar{y} - r\frac{s_y}{s_x}\bar{x}$$

$$b_1 = r\frac{s_y}{s_x}$$

→ rare case! Analytic Formula!

How well does the model predict? — need to know linear model theory

$SSE := \sum_{i=1}^{n} \left(y_i - (b_0 + b_1 x_i)\right)^2$ sum of sqd err.   units: y^2

$MSE := \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - (b_0 + b_1 x_i)\right)^2 = \frac{SSE}{n}$ mean sqd err.   units: y^2 corrected variance of the e's

$RMSE := \sqrt{MSE}$ units: y just like std. dev. has units of supp of r.v., the RMSE has same units

Common to report prediction error as RMSE.

Another very well known metric is called R^2 := "the prop. of variance explained"

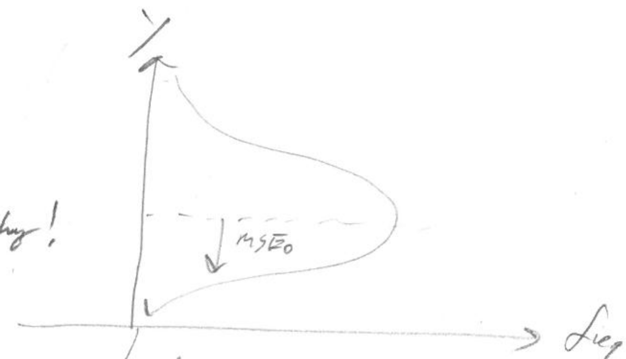
Let's talk about this..

Consider there is no model (i.e. the null model). What is the SSE of this model?

"SST" in textbooks

||

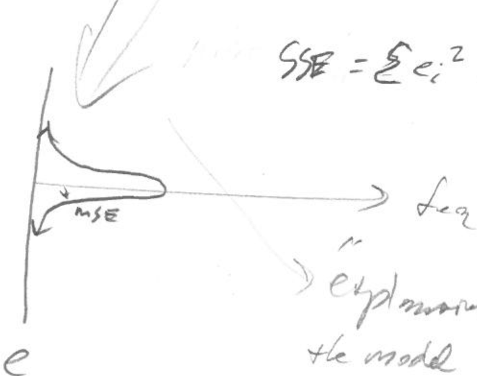
$SSE_0 = \sum (y_i - \bar{y})^2 = (n-1) s_y^2$ exactly the same thing!



After we fit the model, there is a new, hopefully lower SSE.

$SSE = \sum e_i^2 = (n-1) s_e^2$

⇒



"explanation of variance" the model has less error now

How much SSE has been reduced as a proportion of the null SSE?

or // // s^2 // // ... s^2?

$R^2 := \dfrac{SSE_0 - SSE}{SSE_0} = \left.\dfrac{s_y^2 - s_e^2}{s_y^2}\right\}$ prop. of sample variance explained

FYI: $\dfrac{SST - SSE}{SST} = \dfrac{SSR}{SST} = 1 - \dfrac{SSE}{SST}$

estimates... $\approx \dfrac{Var[Y] - Var[E]}{Var[Y]}$

Since $s^2 > 0 \Rightarrow R^2 \leq 1$

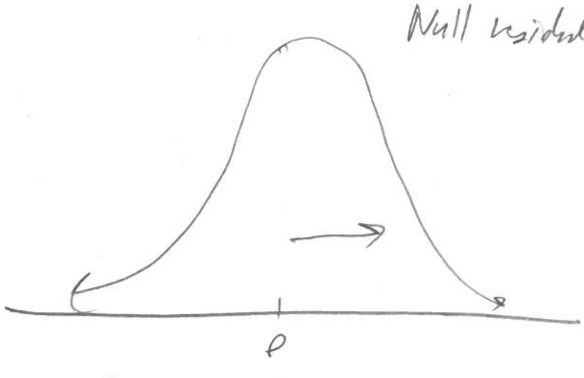
Can $R^2 < 0$? Yes... when if $s_e^2 > s_y^2$ What does this mean? Model is **worse** than null model. It happens!

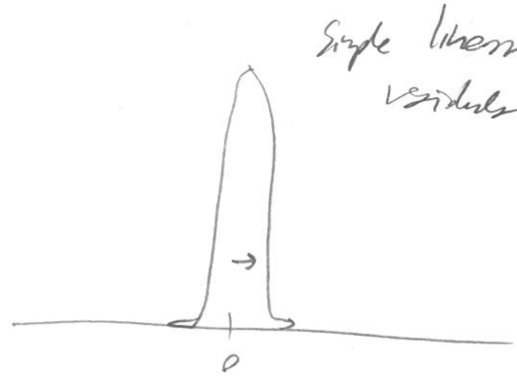
Another way to see this

Residuals

Null model $g(x) = \bar{y}$ $\quad e_i = y_i - \bar{y}$

Single linear regression model $g(x) = b_0 + b_1 x$ $\quad e_i = y_i - (b_0 + b_1 x_i)$





s_e^2 dropped a lot! What % has it dropped? R^2

Another way to assess: