Math 390 Lec 6 2/19/19

1

$\mathcal{Y} = \mathbb{R}$, $\mathcal{H} = \{\vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1}\}$ for the case of $p=1$, $\mathcal{H} = \{w_0 + w_1 x : w_0, w_1 \in \mathbb{R}\}$

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

$g(x) = b_0 + b_1 x$ if $\mathcal{A}$: OLS (ord. least squares) then $b_0 = \bar{y} - r\frac{s_y}{s_x}\bar{x}$, $b_1 = r\frac{s_y}{s_x}$

How well does g predict? "Model performance metric"

$SSE = \sum e_i^2 = \sum_{i=1}^{n} (y_i - g(x_i))^2$ interpretable? units?

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

need to take a class in linear model theory to understand why $n-2$

$RMSE = \sqrt{MSE}$

" ? " ?

just like std dev : variance, RMSE has same units

the model predicts $\pm$ RMSE

$\approx 95\%$ interval for prediction is

$g(x) \pm 2\, RMSE$

How?

Another well known model perf. metric is R^2;

the "prop. of variance explained",

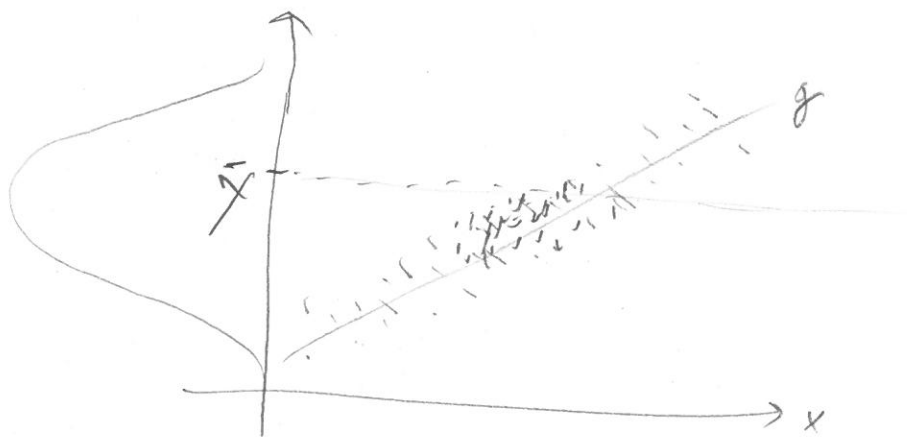
Consider the null model $g_0 = \bar{y}$

What is the SSE of this model?

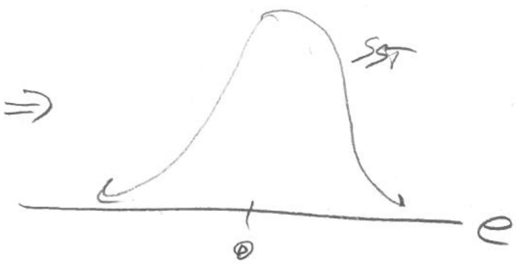
"sum square total"

$$SSE_0 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = SST = (n-1)\, s_y^2$$

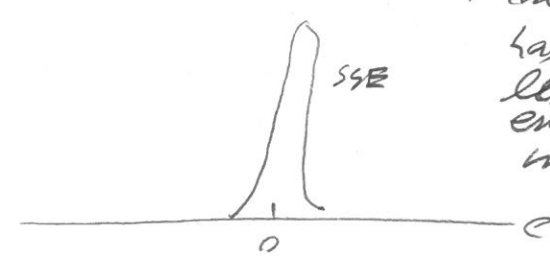
Remember... if you can't beat SST, your model is a loser!



$\Rightarrow$



After model...



model has less error now

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

If I started with s_y^2 as my variance of error in g_0 and now I have s_e^2 ,,,,,,,,,, in g model

then $\Delta s^2 = s_y^2 - s_e^2$ is my reduced variance i.e. the "variance explained"

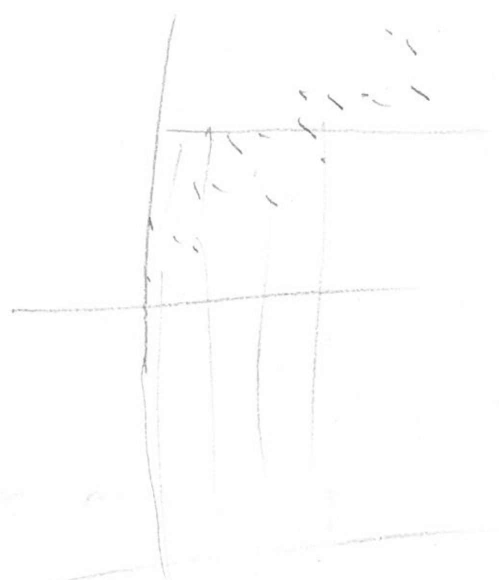
$$\Rightarrow R^2 = \frac{\Delta s^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = \frac{SST - SSE}{SST} \approx \frac{Var[Y] - Var[E]}{Var[Y]}$$

is "sample" prop. of variance explained $\rightarrow$ it's a r.v.!

Can SSE < 0? No $\sum e_i^2 \geq 0$

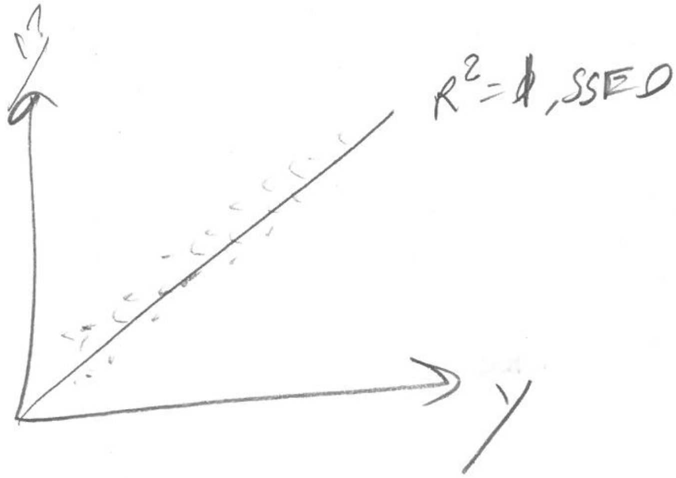
$$\Rightarrow \frac{SST - SSE}{SST} \leq \frac{SST}{SST} = 1 \Rightarrow R^2 \leq 1$$

Can $R^2 < 0$? Yes! A model worse than $\bar{y}$!



$\hat{y}$ terrible $\Rightarrow R^2 < 0$

Another visual. Good for $p > 1$ too.



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

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

RMSE vs. R^2: who is more important?

RMSE! It's in the units you care about!

$R^2 = 99\%$ but RMSE can still be large enough that your model ~~is~~ is good enough.

empirical rule / heuristic (rule of thumb)

$[\hat{y} \pm 2 \cdot RMSE]$ is a 95% confidence set for a prediction

This assumes... prediction errors are normally distributed around $g(\vec{x})$.

---

ex: Special case $p = 1$ but feature is binary.

$X_{raw} \in \mathcal{X} = \{red, green\} \Rightarrow x = \mathbb{1}_{green} \in \{0, 1\}$

$red \downarrow 0$, $green \downarrow 1$ — arbitrary

$\mathbb{1}_{green}$: dummy variable

$\mathcal{H} = \{w_0 + w_1 x : w_0, w_1 \in \mathbb{R}\}$ same!

$\hat{y} = b_0 + b_1 x$

What do you think this model will be?

$$\hat{y} = \begin{cases} \bar{y}_{green} & \text{if } X_{raw} = green \\ \bar{y}_{red} & \text{if } X_{raw} = red \end{cases}$$