2/22/21

$\hat{y} = g(x) = \underbrace{\bar{y}_{red}}_{b_0} + \underbrace{(\bar{y}_{green} - \bar{y}_{red})}_{b_1} x$, let $n_g = \sum x_i$; $p_g = \bar{x} = \frac{n_g}{n}$

$n_r = n - n_g$

$\bar{y} = \frac{1}{n}(\sum y_i) = \frac{1}{n}(\sum_{i:green} y_i + \sum_{i:red} y_i) = \frac{\sum_G y_i}{n} \cdot \frac{n_g}{n_g} + \frac{\sum_r y_i}{n} \frac{n_r}{n_r}$

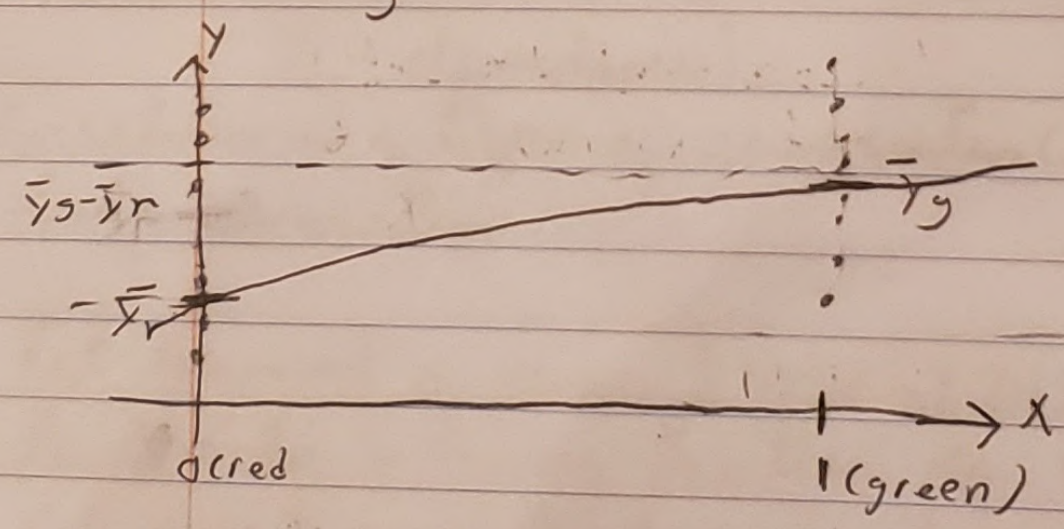
$= p_g \frac{\sum_G y_i}{n_g} + (1-p_g) \frac{\sum_R y_i}{n_r} = p_g \bar{y}_g + (1-p_g)\bar{y}_r$

$n_g \bar{y}_g = \sum_G y_i$ $\quad p_g$

$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\underbrace{\sum x_i^2}_{n_g} - n\underbrace{\bar{x}^2}_{p_g}} = \frac{n_g \bar{y}_g - n p_g \bar{y}}{n_g - n p_g^2} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} = \frac{p_g \bar{y}_g - p_g \bar{y}}{p_g - p_g^2} = \frac{\bar{y}_g - \bar{y}}{1-p}$

$= \frac{\bar{y}_g - (p_g \bar{y}_g + (1-p_g)\bar{y}_r)}{1-p_r} = \frac{(1-p_g)\bar{y}_g - (1-p_g)\bar{y}_r}{1-p_g} = \bar{y}_g - \bar{y}_r$

$b_0 = \bar{y} - b_1 \underbrace{\bar{x}}_{p_g} = p_g \bar{y}_g + (1-p_g)\bar{y}_r - (\bar{y}_g - \bar{y}_r) p_g = \bar{y}_r$



What if $x \in \{red, green, blue\}$? This is then $p=2$ and we need and OLS solution for $p>1$. But intuitively...

$$g(x) = \begin{cases} \bar{y}_{red} \text{ if } x = red \\ \bar{y}_{green} \text{ if } x = green \\ \bar{y}_{blue} \text{ if } x = blue \end{cases} = \underbrace{\bar{y}_{red}}_{b_0} + \underbrace{(\bar{y}_{green} - \bar{y}_{red})}_{b_1} \mathbb{1}_{x=green} x_1 + \underbrace{(\bar{y}_{blue} - \bar{y}_{red})}_{b_2} \mathbb{1}_{x=blue} x_2$$

How well does g predict? We need a "model performance metric". In the SVM this was accuracy or misclassification error. Here, it will can also be what we use externally in the algorithm;

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

Is SSE interpretable? No, lets take the mean at least, call that mean squared error (MSE);

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

But this is still in the squared unit of the phenomenon so its still uninterpretable, We can take the square root of MSE called root mean squared error (RMSE);

$$s_e = RMSE = \sqrt{\frac{1}{n-2} \sum e_i^2} = \sqrt{MSE}$$

RMSE is in the same unit as y (just like the standard deviation is in the same unit as the random variable). Also, from the CLT,

$$[g(x) \pm 1.96 \cdot RMSE]$$ (it is a bit like the SD of the residuals s_e)

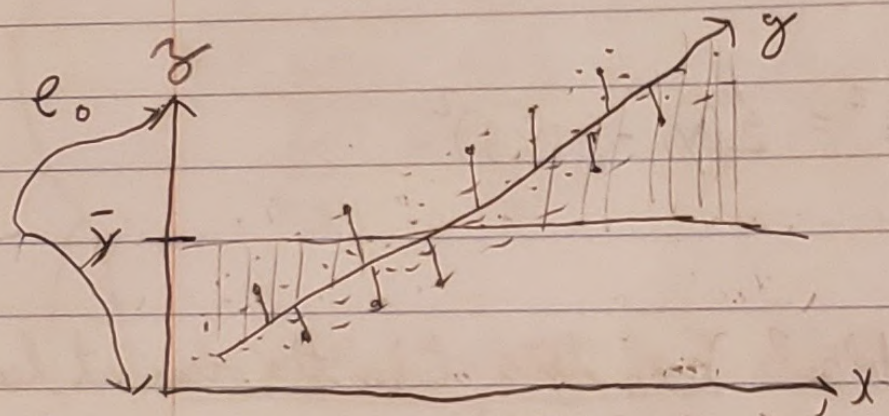
is approx a 95% confidence interval for the true y at that X. RMSE is a very important metric in regression models.

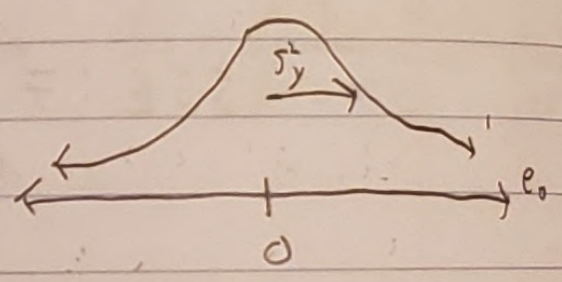
Another important error / performance metric is "R-squared" which is the "proportion of variance explained", we will now explain this definition

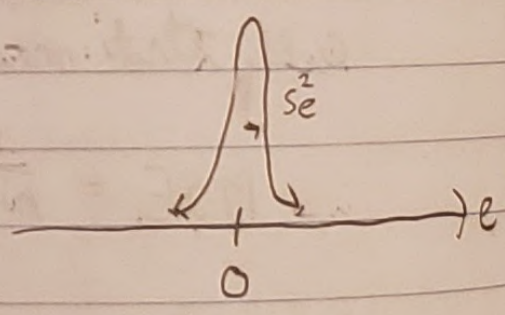
Consider the null model, $g_0 = \bar{y}$. what is the SSE of this model? Let's call it SSE_0.

$$SSE_0 = \sum_{i=n}^{n} e_{0,i}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \underline{SST} = (n-1)s_y^2$$

sum of squares total







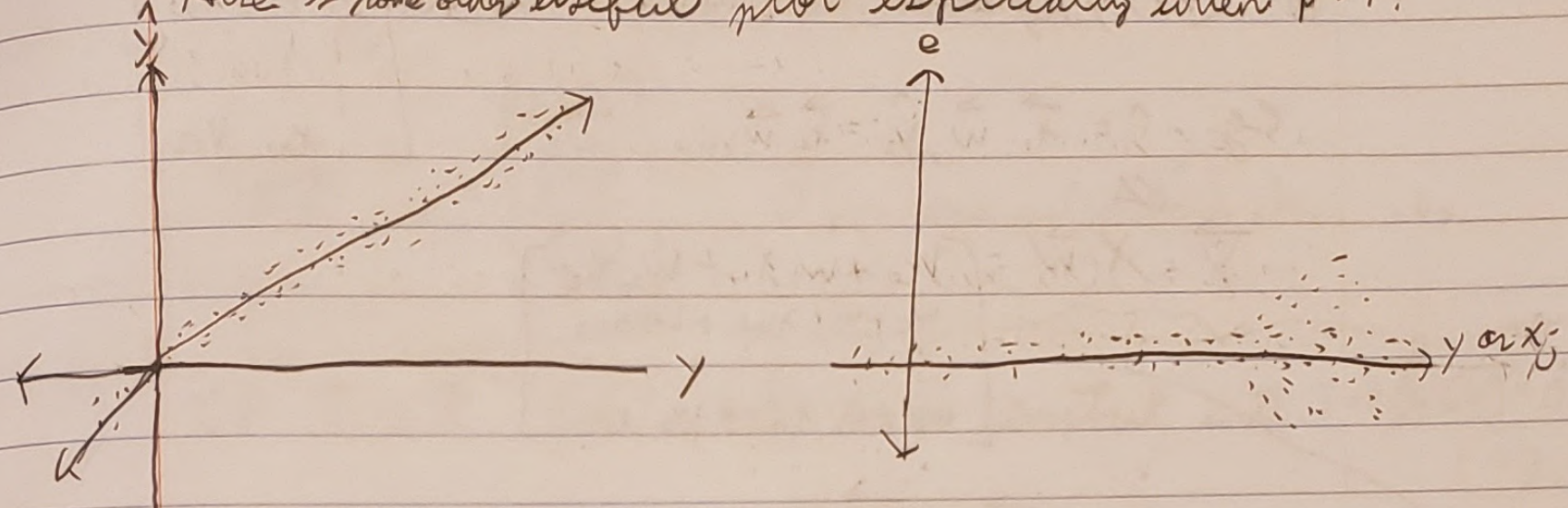
$$\frac{SSE}{SST} = \frac{(n-1)s_e^2}{(n-1)s_y^2} = \frac{s_e^2}{s_y^2}$$

$$R^2 = \frac{SST - SSE}{SST} = \frac{(n-1)s_y^2 - (n-1)s_e^2}{(n-1)s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = \frac{\Delta s^2}{s_y^2}$$

Δs^2

R-squared can never be more than 100%. But R-squared can be negative. This occurs when $s_e^2 > s_y^2$ meaning the model is predicting worse than $g_0 = \bar{y}$

Here's some other useful plot especially when $p > 1$:



$R^2 = 1 \iff RMSE = 0$

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

If $R^2 = 99\%$, does this mean the model is for sure "good"?
No, because if the initial variance was so very large, even a 99% reduction wouldn't result in a small residual variance ie. RMSE still could be high after 99% variance reduction.

We now would like to generalize the least squares estimation algorithm to cases where $p > 1$. Lets begin with $p = 2$.

$\vec{w}$ $\vec{x}$

$$\mathcal{H} = \{ w_0 + w_1 x_1 + w_2 x_2 : w_0, w_1, w_2 \in \mathbb{R} \}$$

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - w_0 - w_1 x_{1,i} - w_2 x_{2,i})^2$$

$$b_0 = \underset{w_0 \in \mathbb{R}}{\operatorname{argmin}} \{SSE\}, \quad b_1 = \underset{w_1 \in \mathbb{R}}{\operatorname{argmin}} \{SSE\}, \quad b_2 = \underset{w_2 \in \mathbb{R}}{\operatorname{argmin}} \{SSE\}$$

This problem can be solved more simply with matrix algebra and a matrix equation:

$D = \langle X, \vec{y} \rangle$, let $X = [\vec{1}_n \ \vec{X}_{\cdot 1} \ \vec{X}_{\cdot 2}] = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{33} \\ 1 & \vdots & \vdots \\ \vdots & & \\ 1 & x_{n1} & x_{n2} \end{bmatrix}$

eg. $\hat{y}_1 = \vec{x}_1 \vec{w}$, $\hat{y}_2 = \vec{x}_2 \vec{w}$, ...

$\vec{\hat{y}} = X\vec{w} = \begin{bmatrix} w_0 + w_1 x_{11} + w_2 x_{12} \\ w_0 + w_1 x_{21} + w_2 x_{22} \\ \vdots \\ w_0 + w_1 x_{n1} + w_2 x_{n2} \end{bmatrix}$

define $\vec{e} := \vec{y} - \vec{\hat{y}}$

$SSE = \sum_{i=1}^{n} e_i^2 = \vec{e}^T \vec{e} = (\vec{y} - \vec{\hat{y}})^T (\vec{y} - \vec{\hat{y}})$

$\vec{\hat{y}}^T \vec{\hat{y}} = (\vec{y}^T - \vec{\hat{y}}^T)(\vec{y} - \vec{\hat{y}})$

$= \vec{y}^T \vec{y} - \vec{\hat{y}}^T \vec{y} - \vec{y}^T \vec{\hat{y}} + \vec{\hat{y}}^T \vec{\hat{y}} = \vec{y}^T \vec{y} - 2\vec{\hat{y}}^T \vec{y} + \vec{\hat{y}}^T \vec{\hat{y}}$

$= \vec{y}^T \vec{y} - 2(X\vec{w})^T \vec{y} + (X\vec{w})^T X\vec{w} = \vec{y}^T \vec{y} - 2\vec{w}^T X^T \vec{y} + \vec{w}^T X^T X \vec{w}$