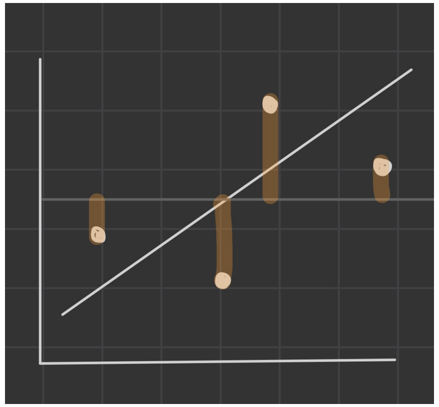
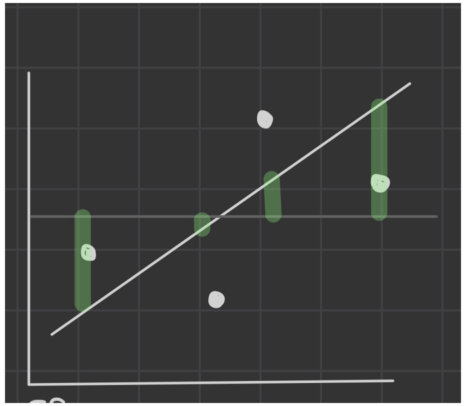
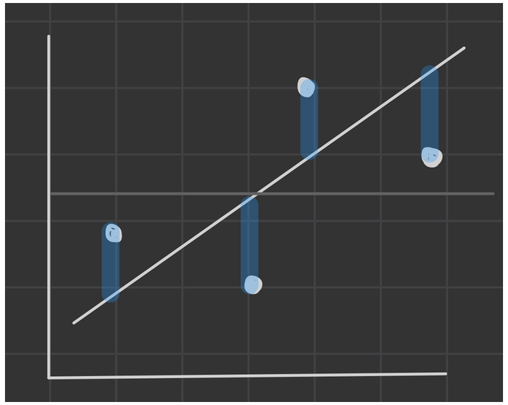
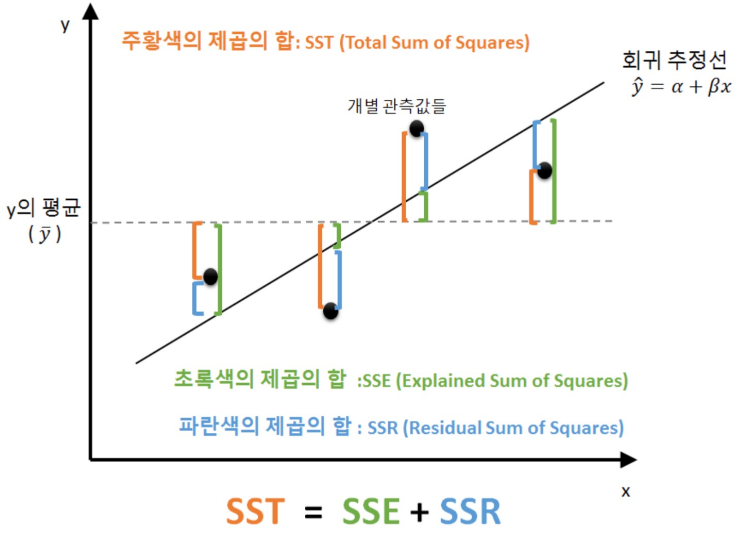
Understanding Evaluation Metrics for regression models

Yay

Which distance?

- Whether absolute or relative; **distance metric**
- **SSR**: Residual is what **absolute metrics** are interested about.
- **SST**: Mean of actual y variables are set as the ‘baseline’ in order to achieve the **relativeness.**

Total Sum of Squares (SST)	Explained Sum of Squares (SSE)	Residual Sum of Squares (SSR)
$\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$	$\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$
		



Absolute vs. Relative?

- How **accurate** is the model? vs. To **which degree** does the model explain?
- Aims **lower** vs. aims **higher**

Absolute			Relative	
0 to infinity			0 -1	0 -1 (always $\leq R^2$)
Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	R-square (Coefficient of determination)	Adjusted R-squared
$\frac{1}{n}\sum\lvert y_j - \hat{y}_j\rvert$	$\frac{1}{n}\sum(y_j - \hat{y}_j)^2$	$\sqrt{\frac{1}{n}\sum(y_j - \hat{y}_j)^2}$	$1 - \frac{\sum(y_j - \hat{y}_j)^2}{\sum(y_j - \overline{y}_j)^2}$	$1 - \frac{(1-R^2)(n-1)}{n-k-1}$

Absolute in detail : MAE vs. MSE vs. RMSE

- Common confusion in metrics
- MAE derived the first, MSE derived in support (Actual usage depends on the dataset)
- Differentiability allows model optimization (i.e. Gradient descent)

Mean Absolute Error (MAE)	**Mean Squared Error (MSE)**	**Root Mean Squared Error (RMSE)**
Average magnitude of errors Average of residuals	Average of squared errors Average of squared residuals	Square root of average of squared errors
$\frac{1}{n}\sum\lvert y_j - \widehat{y_j}\rvert$	$\frac{1}{n}\sum(y_j - \widehat{y_j})^2$	$\sqrt{\frac{1}{n}\sum(y_j - \widehat{y_j})^2}$
*Mean Bias Error(MBE) $\frac{1}{n}\sum(y_j - \widehat{y_j})$	*MSE ≠ Variance(y) $\frac{1}{n}\sum(y_j - \overline{y_j})^2$	*RMSE ≠ Standard Deviation(y) $\sqrt{\frac{1}{n}\sum(y_j - \overline{y_j})^2}$
Intentionally canceling out positive and negative errors to measure model bias	If it is the MSE of a base model based on the average y true values (i.e. LR(x) = $\bar{y}$), Variance(y) would be the same. However, the difference is how **MSE focuses on the variability around the predicted regression line whereas the variance sees the variability around a horizontal line of a mean value**. This applies equally to the relationship between RMSE and SD(y).	
• MAE assigns identical weight for all residuals. • Robust with outliers. • **Non differentiable.**	• Squaring penalizes the large prediction errors. • Thus, affected by outliers. • Yet, effective if large errors undesirable. • **Differentiable.**	• Square root doesn't prevent squaring from penalizing the large prediction errors. • Thus, affected by outliers. • **Differentiable** • **Has same units as dependent variable**

Absolute in detail : MAE vs. RMSE

- While MAE is steady, RMSE increases as the variance increases. However, RMSE does not necessarily increases with the variance of the errors. **RMSE increase with the variance of the frequency distribution of error magnitude.**
- MAE ≤ RMSE : the RMSE result will always be larger or equal to the MAE. Equal when all of the errors have the same magnitude.
- RMSE ≤ MAE*sqrt(N): When all of the prediction error comes from a single sample, the squared error then equals to (MAE^2)*N.

CASE 1: Evenly distributed errors

ID	Error	|Error|	Error^2
1	2	2	4
2	2	2	4
3	2	2	4
4	2	2	4
5	2	2	4
6	2	2	4
7	2	2	4
8	2	2	4
9	2	2	4
10	2	2	4

MAE	RMSE
2.000	2.000

CASE 2: Small variance in errors

ID	Error	|Error|	Error^2
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	3	3	9
7	3	3	9
8	3	3	9
9	3	3	9
10	3	3	9

MAE	RMSE
2.000	2.236

CASE 3: Large error outlier

ID	Error	|Error|	Error^2
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	20	20	400

MAE	RMSE
2.000	6.325

CASE 4: Errors = 0 or 5

ID	Error	|Error|	Error^2
1	5	5	25
2	0	0	0
3	5	5	25
4	0	0	0
5	5	5	25
6	0	0	0
7	5	5	25
8	0	0	0
9	5	5	25
10	0	0	0

var	MAE	RMSE
6.944	2.500	3.536

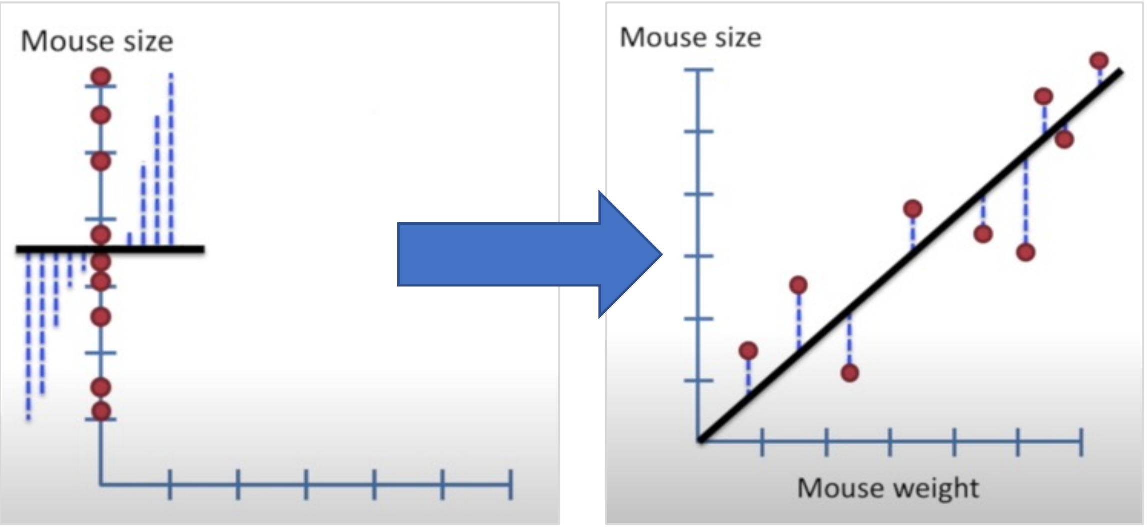
CASE 5: Errors = 3 or 4

ID	Error	|Error|	Error^2
1	3	3	9
2	4	4	16
3	3	3	9
4	4	4	16
5	3	3	9
6	4	4	16
7	3	3	9
8	4	4	16
9	3	3	9
10	4	4	16

var	MAE	RMSE
0.278	3.500	3.536

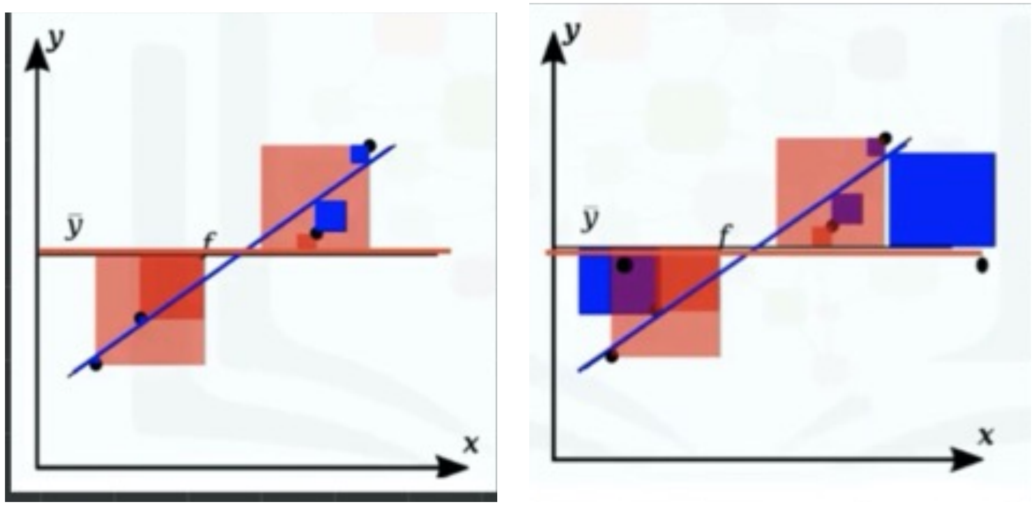
Relative in detail

- Compare the portion of SST errors reduced, by taking independent variables into accounts.
- Think in terms of Variance as ”Sum of Squares” are what remains in calculation.



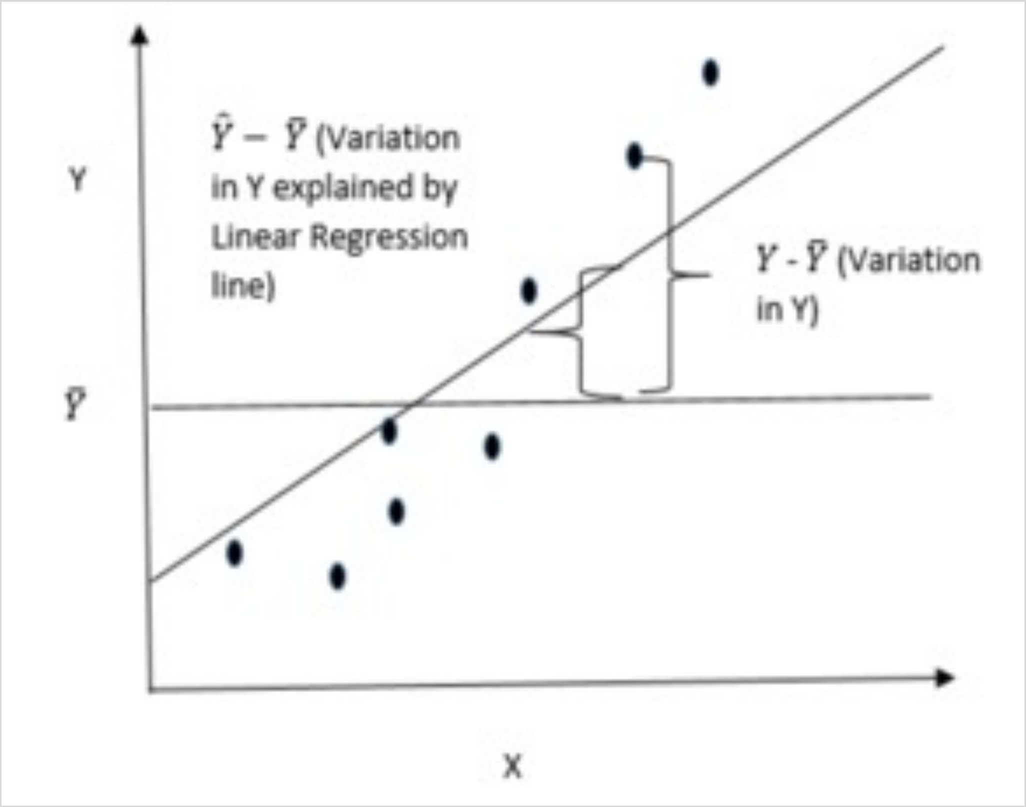
$$R^2 = \frac{\mathrm{Var(mean)} - \mathrm{Var(fit)}}{\mathrm{Var(mean)}} = \frac{\text{Variation in target explained by feature}}{\text{Variation in target withoiut taking any feature into account.}}$$

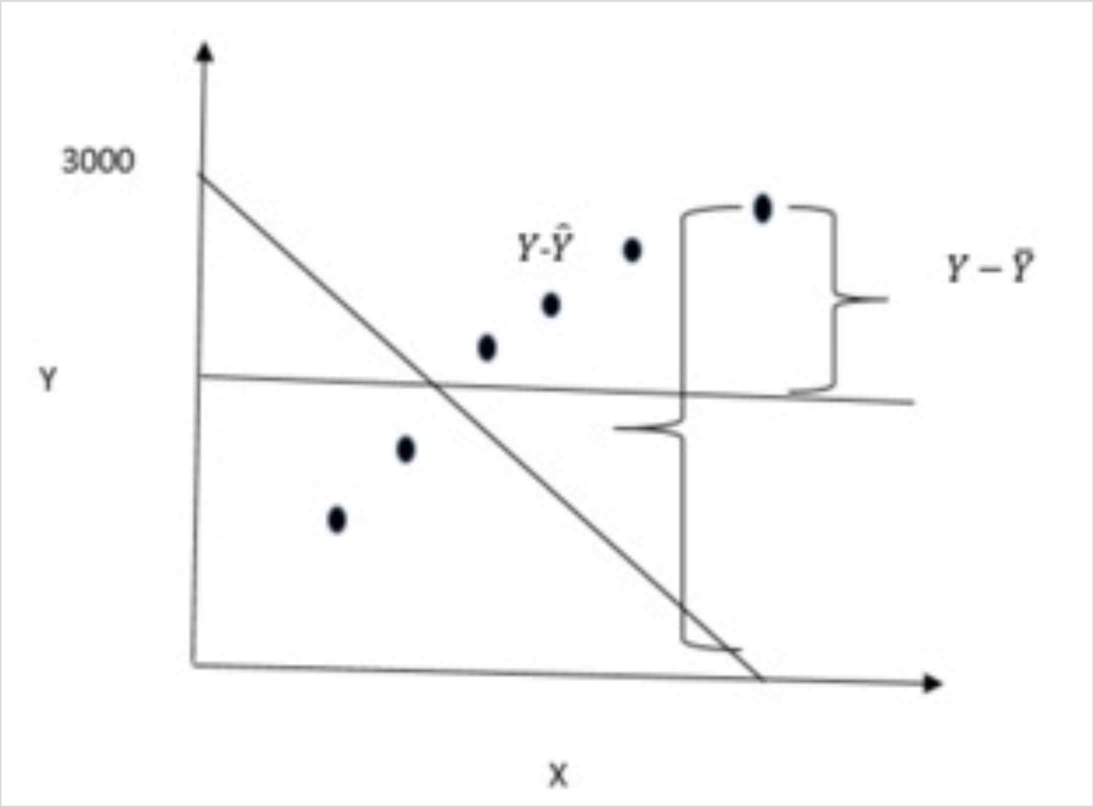
$$R^2 = \frac{\mathrm{SS(mean)} - \mathrm{SS(fit)}}{\mathrm{SS(mean)}}$$



Relative in detail : Negative R-Square

- The term “Square” often gives impression that R-square cannot be less than 0.
- Simply put, the model is clearly “underfitting”, worse than the base model of “Mean” of y variable.





Relative in detail : R-Square vs. Adjusted R-Square

- Compare the portion of SST errors reduced, by taking independent variables into accounts,

R-square (Coefficient of determination)	Adjusted R-squared
0 -1	**0 -1 (always $\leq R^2$)**
1 - $\frac{\sum(y_j-\widehat{y_j})^2}{\sum(y_j-\overline{y_j})^2}$	**1-**$\frac{(1-R^2)(n-1)}{n-k-1}$
$\frac{SST-SSE}{SST}$=1 - $\frac{SSE}{SST}$ = $\frac{SSR}{SST}$ *proportion of variation in the y-variable that is due to variation in the x-variables	Adjusted for the number of independent variables *n : number of observations in the data *k : number of independent variables in the data
R Squared value always increases with the addition of the independent variables which mi ght lead to the addition of the redundant variables in our model. However, the adjusted R -squared solves this problem.	The value of Adjusted R squared decreases if the increase in the R square by the additiona l variable isn't significant enough. *penalizes having a large number of parameters

References

- [Link] What are RMSE and MAE?, towardsdatascience
- [Link] MAE and RMSE – Which Metric is Better?, Medium
- [Link] MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared – Which Metric is Better, Analytics Vidhya
- [Link] RMSE or MAE? Arguments against avoiding RMSE in the literature, ResearchGate

- [Link] Art of Choosing Metrics in Supervised Models Part 1, towardsdatascience
- [Link] Which Evaluation Metric Should You Use in Machine Learning Regression Problems? – towardsdatascience
- [Link] 회귀분석 결과에서 SST, SSE, SSR, R-Squared의 의미, igija tistory

- [Link] When is R Squared negative?, Cross Validated
- [Link] Ambiguities inherent in sums-of-squares based error statistics, ResearchGate

EOD

Yay