Feature Scaling
29 July 2025 22:36

Is feature scaling a must everytime we perform linear regression ?

No!!

When ?

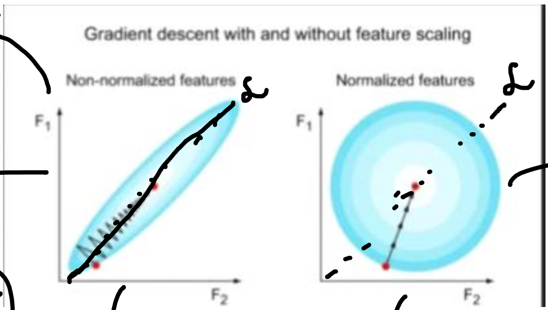
i) Interpretability with respect to Feature importance → if Features are in different scale

ii) Gradient descent {converges Faster} → Why ?

converges → Solution is found faster

Longer

different range (age vs height)



Shorter

elliptical
Non-Uniform

Uniform

Evaluation Metrics (in case of doubts)

30 July 2025 19:15

y_true	y_pred	Error (y_pred - y_true)	Absolute Error	Squared Error
3.0	2.5	-0.5	0.5	0.25
4.5	5.0	0.5	0.5	0.25
2.0	2.4	0.4	0.4	0.16
5.5	5.1	-0.4	0.4	0.16
6.0	6.0	0.0	0.0	0.00

- **Mean Error (ME)** = 0.0 / 5 = **0.0**
- **MAE** = (0.5 + 0.5 + 0.4 + 0.4 + 0.0) / 5 = **0.36**
- **MSE** = (0.25 + 0.25 + 0.16 + 0.16 + 0.0) / 5 = **0.164**

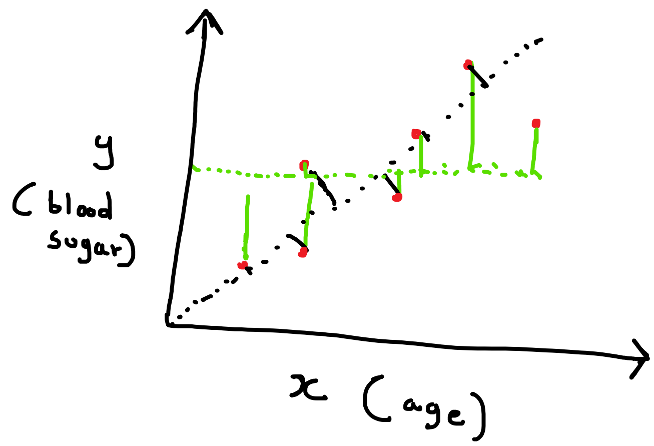
Absolute error is not differentiable at some points $\{ y = y_{pred} \}$

	x	y	y_pred	mean_of_actual_y	tss_for_datapoint (square of (actual_y-mean_of_actual_y))	rss_for_datapoint (square of (actual_y - pred_y))
0	1	2	2.8	4.0	4.0	0.64
1	2	4	3.4	4.0	0.0	0.36
2	3	5	4.0	4.0	1.0	1.00
3	4	4	4.6	4.0	0.0	0.36
4	5	5	5.2	4.0	1.0	0.04

R squared

29 July 2025 23:00

Training data



Mean = $\bar{y}$ => target variable mean

$$TSS = \sum (y_i - \bar{y})^2$$

Total Variation in Value of y (target)

|| Amount of info

$$RSS = \sum (y_i - \hat{y})^2$$

after line is fit

squared difference between predicted and actual

$$R^2 = 1 - \frac{RSS}{TSS}$$

||

$$1 - \frac{\text{Model can't explain}}{\text{Total information in data}}$$

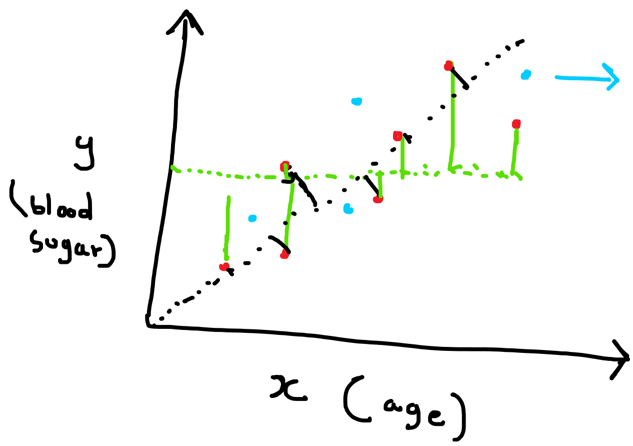
||

= 1 - Proportion of info not explained by model

= Proportion of info/Variance that model was able to explain

Total error => how much Variation the best fit line (Model) could not explain

80% to train 20% to test

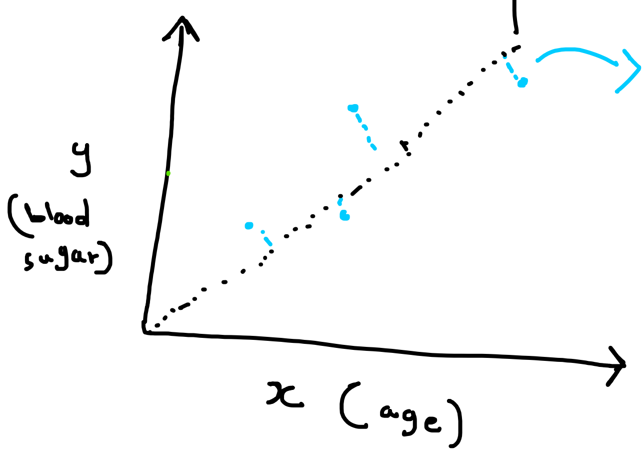


Test data => Why do we need?

How will my model perform on unseen data

Let us seperate test data from train data?

line is purely based on training data



Testing R^2 is calculated

$$1 - \frac{\text{Rss of test data}}{\text{Tss of test data}}$$

Test R^2 gives a more realistic evaluation of the model!!

Train R^2 vs Test R^2 → how much variance unseen does it capture

Training data used for?

Testing data used for?

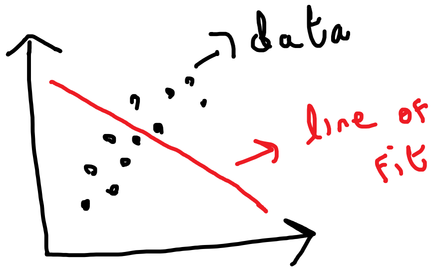
Examination Analogy

Range of R^2? $-\infty$ to 1

$$R^2 = 1 - \frac{RSS}{TSS} > 1$$

Can it be negative? Yes!!

$\frac{RSS}{TSS} > 1$ or $RSS > TSS$



Can it be greater than 1?

No ⇒ You cannot capture more information is already there

$$1.2 = 1 - \frac{RSS}{TSS}$$

$$1.2 - 1 = -\frac{RSS}{TSS}$$

$$0.2 = -\frac{RSS}{TSS} \qquad \frac{RSS}{TSS} = -0.2$$

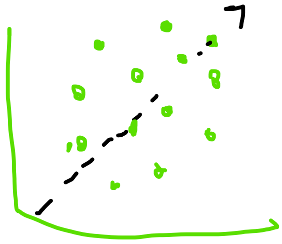
NoT Possible!!

Adjusted R Squared

29 July 2025 23:41

What happens to my {training R^2} if I keep adding variables? => Always stay the same or increase

A model wants to makes full use of all variables/data provided to it.



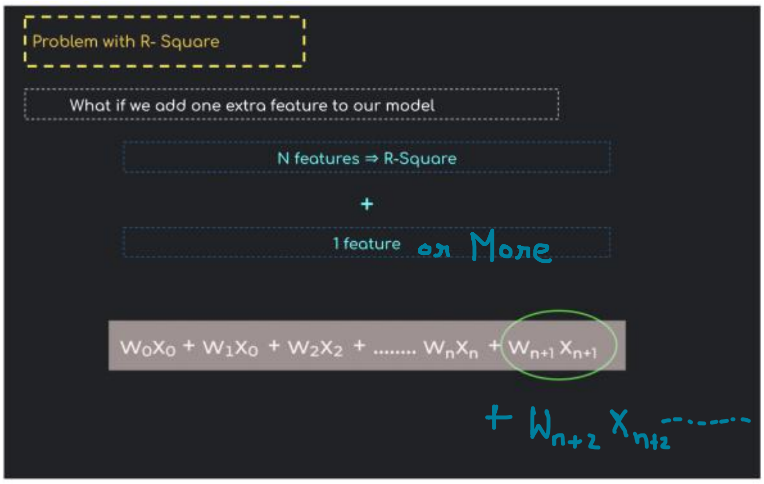
relationship is forced

How about test R^2?

↳ Can move in any direction on addition of variables

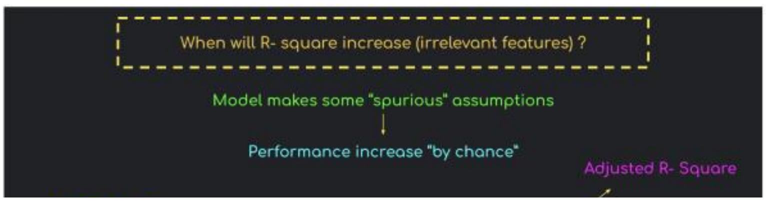
Sometimes we dont have enough data to have train and test data. I have only 50 datapoints

Solution? => Adjusted R-squared



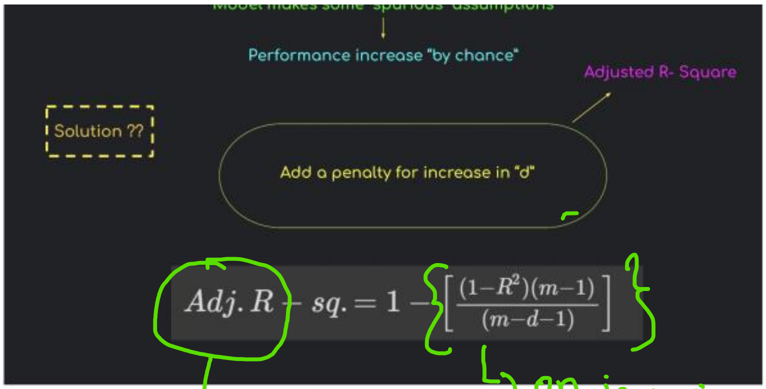
Train $R^2 \geq$

Without testing data, how do I know if the increase is significant?



↳ Sample size

n - No. of



reduce

on increasing d, reduce

m : No. of rows/observations

Sample size

d = No. of features

Adding more features {penalizes R^2} (Adj R^2↑)

If I build 5 different models to predict the same target but with different number of Variables, how will I pick the best Model?

Range of Adj R^2 => $-\infty$ to 1

i) Train v Test R^2 => Access to enough data

ii) Adjusted R^2 => If not

Assumption of Linearity

30 July 2025 16:02