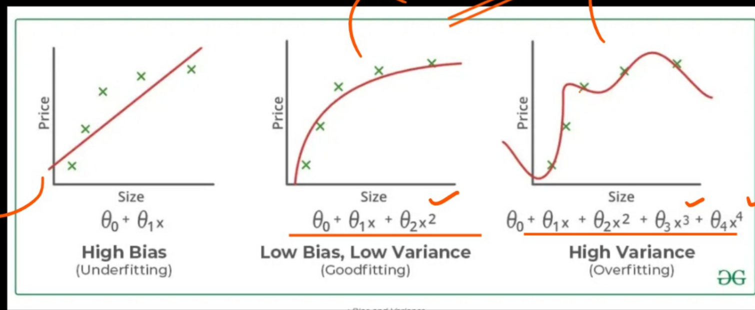
Overfitting & Underfitting

↓ Error = ↓Bias + ↓variance → variation b/w training set vs test set error

↳ training set error

Residual ≠ 0   Residuals = 0

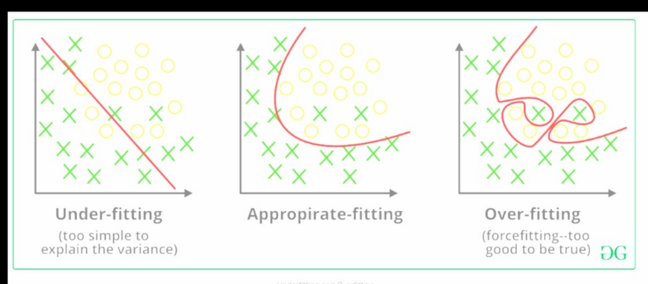
$\hat{y} = mx + c$



underfitting → High Bias + Low variance

Overfitting → Low Bias + High variance

↳ ≈ 0



$\hat{y} = m_1 x_1 + m_2 x_2 + c$

Regularization Techniques → to avoid the issue of overfitting by modifying the cost function of a given model.

↗ feature selection

① Lasso Regression (L1 Regularization)

$$\text{new cost function} = \underbrace{MSE}_{\to 0} + \lambda \sum_{i=1}^{n} |m_i|$$

coefficients

λ → hyperparameter

↳ $|1 + 2| = 3$

$m_i \approx 0$

↳ $i = 1$ to n

↳ the feature which belongs to m_i that is irrelevant

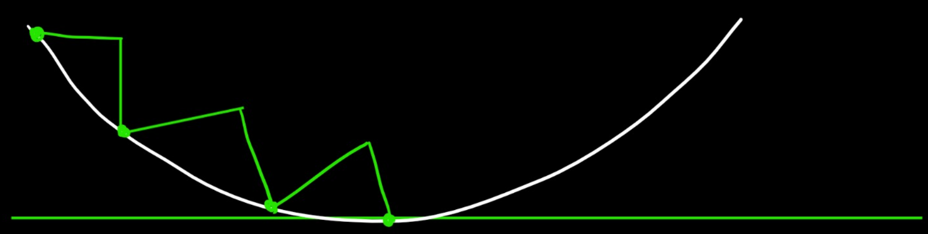
↓ overfitted

$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

② Ridge Regression (L2 Regularization)

$(1^2 + 2^2)$

New cost function $\Longrightarrow MSE + \lambda \sum_{i=1}^{n} m_i^2$



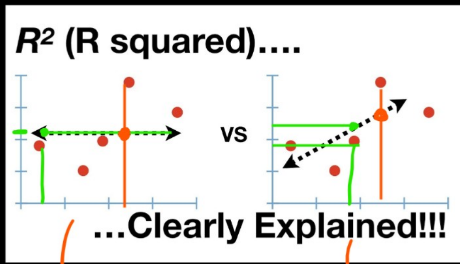
③ Elastic Net Regression $\longrightarrow$ Combination of L1 & L2

New cost function $\Rightarrow MSE + \lambda_1 \sum_{i=1}^{n} |m_i| + \lambda_2 \sum_{i=1}^{n} m_i^2$

alpha $\alpha \longleftarrow$

- $\lambda_1 = 0 \longrightarrow$ Ridge Regression
- $\lambda_2 = 0 \longrightarrow$ Lasso Regression
- $0 < \lambda_1, \lambda_2 < 1 \longrightarrow$ Mix of both Lasso & Ridge

R- squared



Denominator

numerator

Always increases or stays the same

$$\boxed{R^2 = 1 - \frac{SS_{reg}}{SS_{total}}}$$

$$R^2 = 1 - \boxed{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}}$$

Numerator ↓↓ → new predictor

↑Denominator

num > Den → best fit line is not for now good

Increase → 1 - (small value)

num < Den → ideal situation

higher the value of R^2 → better the performance of the model

$(0 < R^2 < 1)$

Adjusted R-square

Increases only if the predictor improves fit, otherwise it should decrease.

new feature → location

↳ improves

new feature → # children

↳

$R^2 \approx 1$ → good fit

$R^2 \approx 0$ → poor fit

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

N Total Sample Size

p Number of independent variable

← how irrelevent features

↳ Adjusted ↓↓ R squared