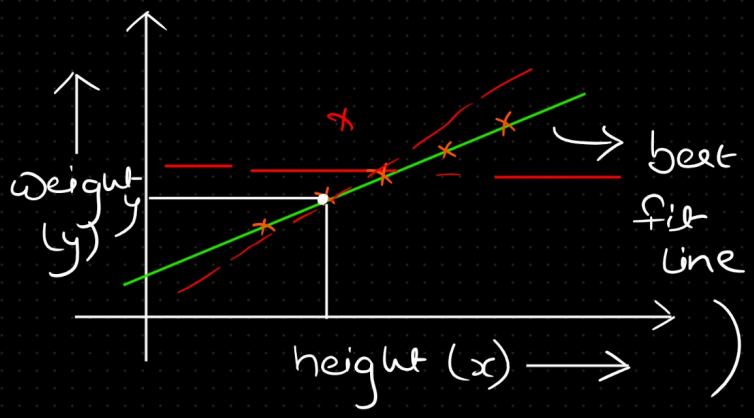
# Regularization Techniques

Overfitting → High variance, Low Bias




train = 99.99% ↑↑
Error ≈ 0

test = 40% ↓↓
Error ≈ ↑↑

Error ≈ 0
train data

## Error

RSS

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

(Linear Regression)

n → num of records

## L2 Regularization (Ridge Regularization)

hyperparameter k → # features

cost function → $RSS + \lambda \sum_{i=1}^{k}(m_i)^2$

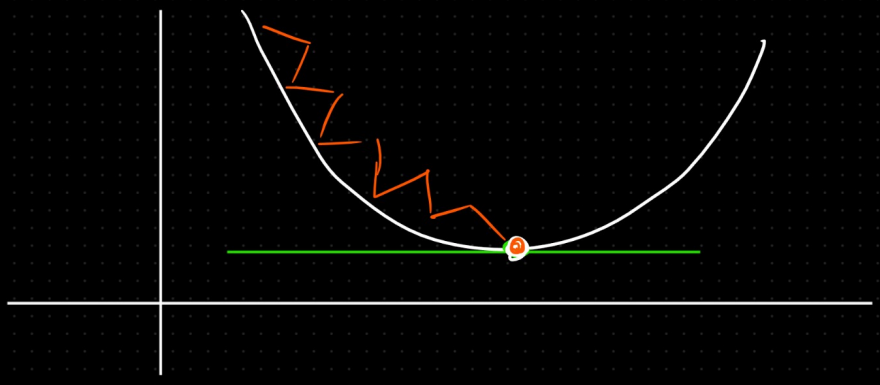
$0 + (>0)$

slope

$y = m_1x + m_2x^2 + m_3x^3 + c$

Intercept

Gradient Decent



L1 Regularization (LASSO)

(feature selection)

Least Absolute Shrinkage & Selection Operator

Cost function → $RSS + \lambda \sum_{i=1}^{n} |m_i|$

$0 + (>0)$

for few feature

$m \approx 0$

Elastic Net

- Reduce overfitting
- feature selection

↓

Combination of L1 & L2

Regularization Techniques

$$\lambda_1 \sum_{i=1}^{k} |m_i| + \lambda_2 \sum_{i=1}^{k} m_i^2$$

L1

L2