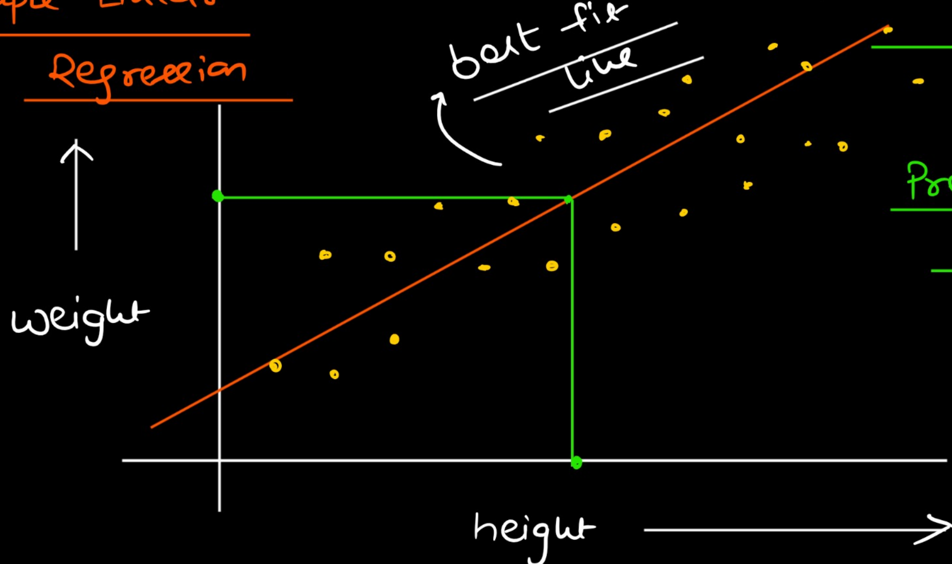
Linear Regression

Supervised Learning → Regression Problems (Predicting Continuous valued outputs)

(Stock Price Prediction, Real Estate Prediction)

Mathematical Intuition Behind Linear Regression

Simple Linear Regression



$\hat{y} = mx + c$

Predicted Values ($\hat{y}$), slope (m), intercept (c), height (x)

$(y - \hat{y}) \longrightarrow$ Error / Residual ↓↓

Real time scenarios ⟶ Multiple Linear Regression

$$\hat{y} = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + c$$

$\{x_1, x_2, x_3, x_4, x_5\} \longrightarrow$ features of the given data

Training Linear Regression Model

↓

Optimal value of coefficients

(m & c)

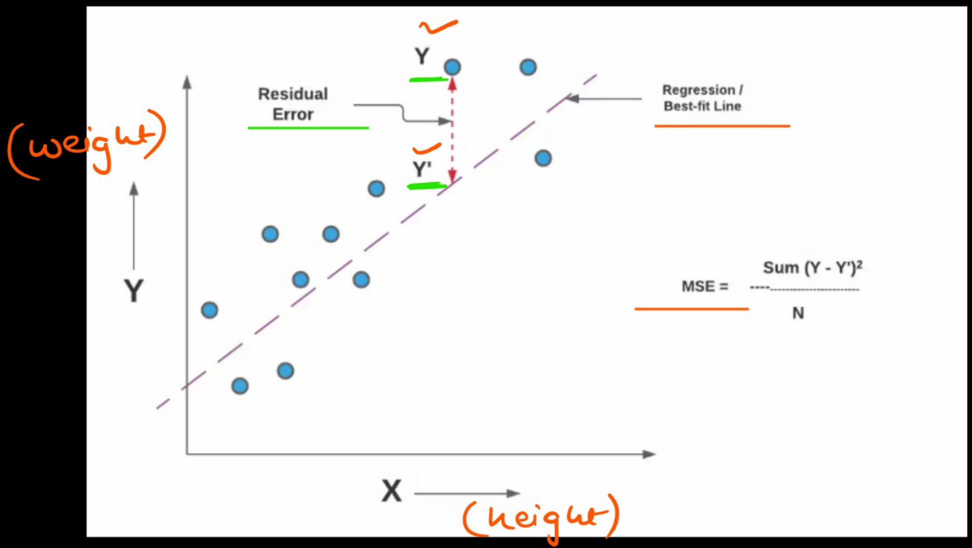
Error → MSE (Mean Squared Error)

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

predicted

actual

$n \rightarrow$ # data points

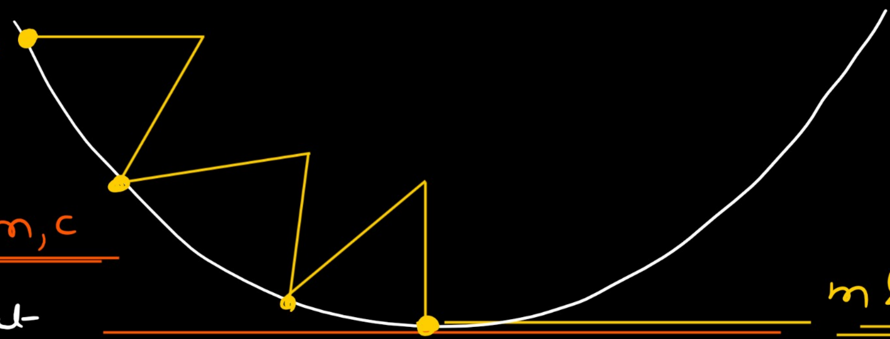


- Best fit line
- Residual Error
- Mean Squared Error

Optimizers → Gradient Descent

↳ to get the optimal value of m & c

m, c



m, c

m & c ??

Gradient Descent

Algorithms

m, c

single local/global

minima

$$\left\{\begin{aligned} m_{new} &= m_{old} - \alpha \frac{\delta L}{\delta m} \\ c_{new} &= c_{old} - \alpha \frac{\delta L}{\delta c} \end{aligned}\right\}$$

~~Rate of change~~

L → Loss function → MSE

vs

Cost function

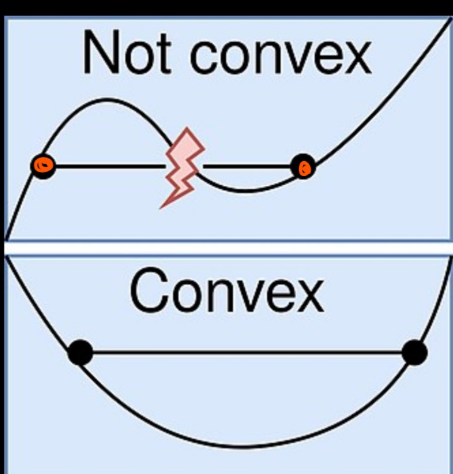
$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Computationally easy

$$\text{Loss function } (L) \rightarrow \frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

$$\Rightarrow \frac{1}{n}\sum\left(y_i - (mx_i + c)\right)^2$$

$$\frac{\delta L}{\delta m} = \frac{1}{n}\sum 2\left(y_i - (mx_i + c)\right)(-x_i)$$

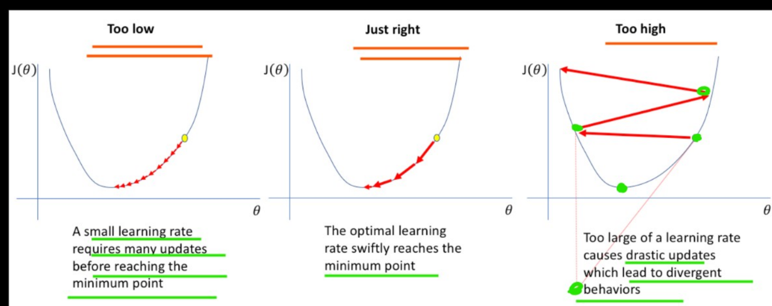


$$\boxed{\frac{\delta L}{\delta m} = -\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - \hat{y}_i)}$$

$$\frac{\delta L}{\delta c} = \frac{1}{n} \Sigma\, 2\left(y_i - (mx_i + c)\right)(-1)$$

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

Learning Rate



Hyperparameter

$\alpha = 0.01$

Predicted value

Loss function vs Cost function

loss function = $(y - \hat{y})^2$

target value

Loss function: measure the error for a single data point

Cost function: average loss across the entire dataset

~~Gradient Descent~~ ← Cost function $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Assumptions of Linear Regression

① Linearity

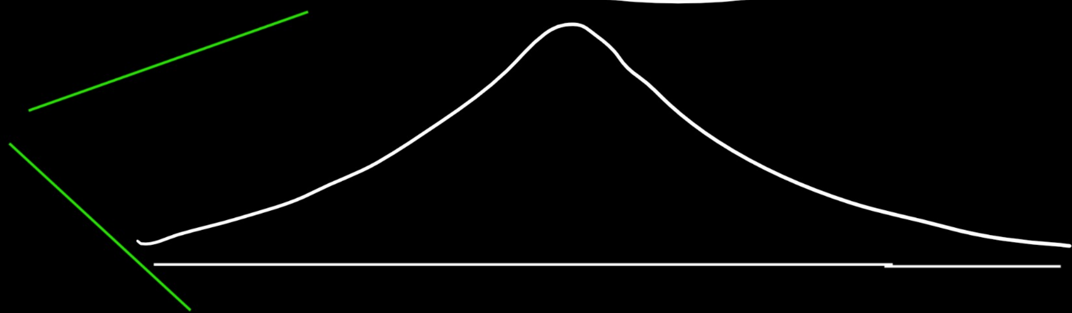
- Independent → x (height)
  $(x_1, x_2, x_3, x_4, x_5)$
- Dependent → $\hat{y}$ (weight)
  - Relationship should be linear in nature

② Independence → $\underbrace{(x_1, x_2, x_3, x_4, x_5)}_{\text{independent of each other}}$

(x_1) Income	(x_2) Expenditure

→ Dependent of each other

3 Normality → Residual Error Distribution



4 Homoscedasticity → Variance of the Residuals

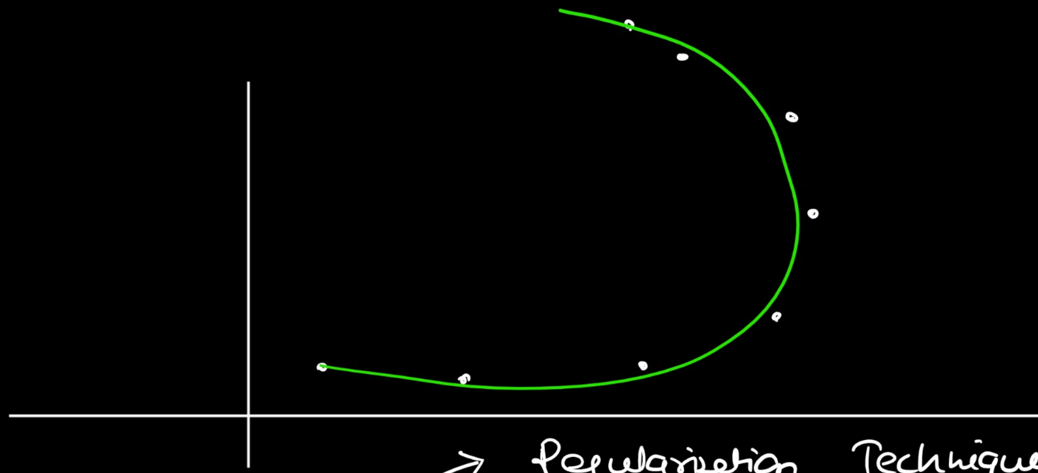
Constant

Statistical Validation

→ Crucial valid hypothesis testing & Confidence intervals

Task 1

Polynomial Regression



Regularization Techniques

↳ Lasso

Ridge

Elastic Net

Overfitting & Underfitting

Implementation of LR

Model Evaluation

RMSE,

MAE,

R1- Score