Simple Linear Regression

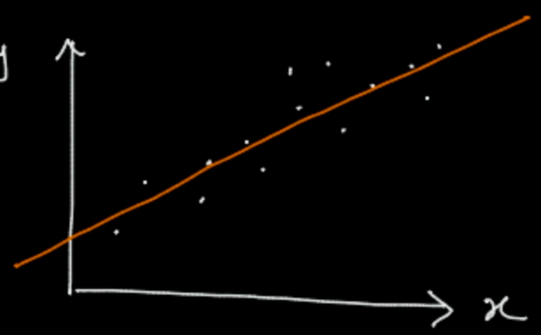
Only one indep Var

Target is continuous.

$$y = f(x)$$

$$\hat{y} = mx + c$$

1. Scatter plot



2. Strength of linear relationship.

Correlation coeff (r) $\Rightarrow -1$ to $+1$

$r \geq \pm 0.85$

Find the line of Best fit → that line can best represent the data

Metrics → MSE

→ RMSE = $\sqrt{\text{MSE}}$

→ MAE

R^2 → % of Variance that can be explained by the model

x	y	$\hat{y}$	$(y - \hat{y})^2$	$\lvert y - \hat{y} \rvert$
			↓	↓
			MSE	MAE

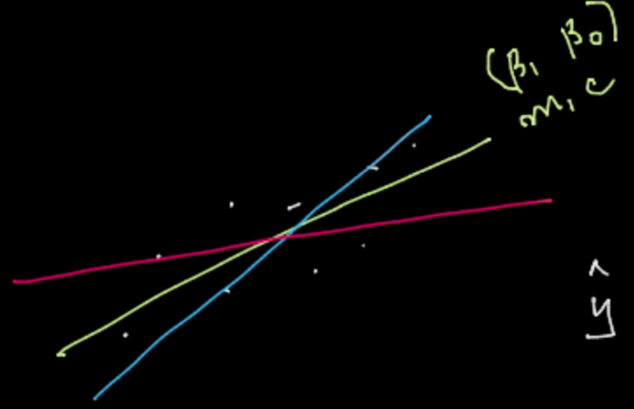
x_1 x_2 y $\hat{y}$

← Multi linear Regression →

$$y = f(x)$$

β_0, β_1

↳ $\hat{y} =$



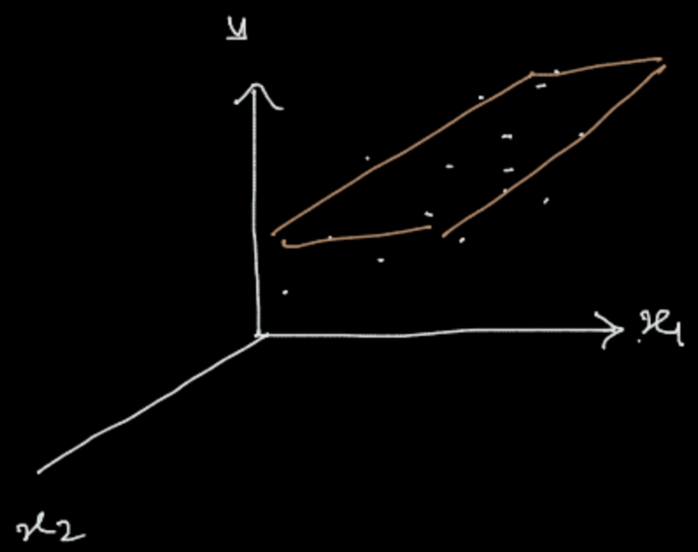
$$\hat{y} = \beta_0 + \beta_1 x \longrightarrow \text{straight line}.$$

$y = mx + c \longrightarrow$ st. line

$y = \beta_0 + \beta_1 x$

↓ ↳ slope (m)

intercept (c)



$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \longrightarrow \text{plane}$$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_m x_m \longrightarrow \text{Hyper plane}$$

$m \rightarrow$ no. of features

x_1 x_2 x_3 . . . . x_m	y	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$
				↓

unstandardised ⟨ - MSE
- RMSE $= \sqrt{\text{MSE}}$

↳ Standardised

$R^2 \rightarrow \geq 85\%$.

1. 0.85 0.76 0.43 — Strong linear rel.

x_1	x_2	x_3	y

$r \approx 0$.

→ Scatter plot

→ Correl coeff

2. The features should be independent of each other.

Otherwise → "Multi collinearity" Problem

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Area, Rent, distance

↑

↳ Model parameters

→ Weights

$x_1 \uparrow$ $x_2 \uparrow$

Model Training → Model extracts the model parameters from the data

→ $\beta_0, \beta_1, \beta_2, \beta_3$

Hold all values Constant

Change only x_2 by 1 unit.

How much change in there in 'y'

→ β_2

$\beta_0 = 2$, $\beta_1 = 3$, $\beta_2 = 4$, $\beta_3 = 5$; $\beta_0 = 100$.

$\uparrow x_1 = 1000$, $x_2 = 2\uparrow$, $x_3 \to 5$ km

$$y_1 = 100 + 3 \times 1000 + 4 \times 2 + 5 \times 5$$
$$= 100 + 3000 + 8 + 25$$
$$= 3133.$$

$$y_2 = 100 + 3 \times 1000 + 4 \times 3 + 5 \times 5$$
$$= 101 + 3000 + 12 + 25$$
$$= 3137.$$

$y_2 - y_1 = 4$.

sklearn.

- linear_model
  - LinearRegression → class
- metrics
  - mean_squared_error() ← Function
  - → r2_score() ←
- model_selection.
- preprocessing
- tree
- ensemble

(actual, predicted) — y, ŷ

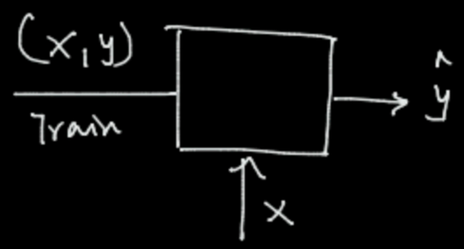
obj = ClassName(!) ← Param

reg_model = LinearRegression( )

- reg_model ↑ Snake case
- LinearRegression → TitleCase

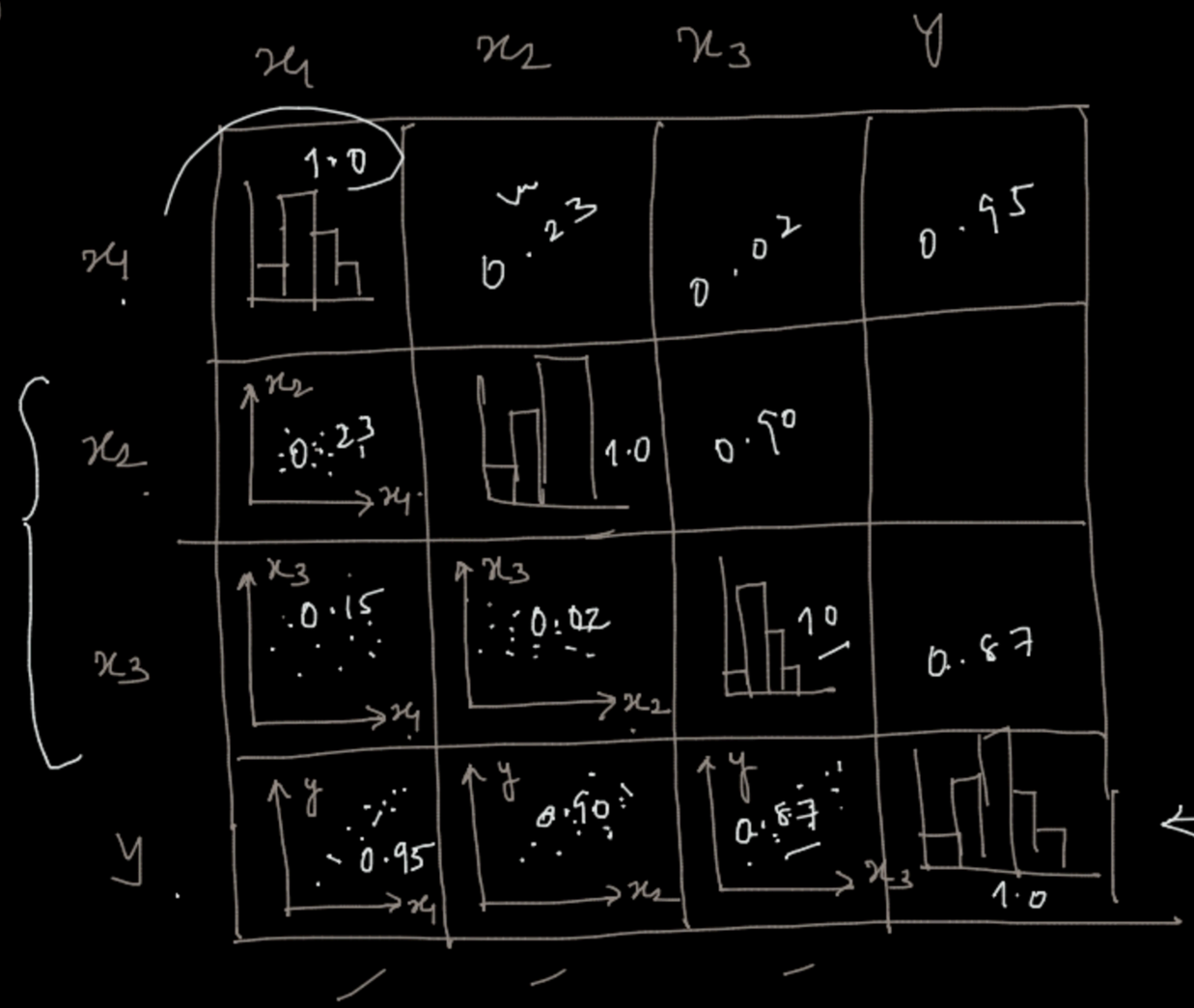
from sklearn.linear_model import LinearRegression

(a) Create the model → ClassName( )
(b) train the model → fit( )
(c) Predict → Predict( )
(d) Metrics Calculation →



Matrix plots

1. pair plot
2. Heat map



Heat map.