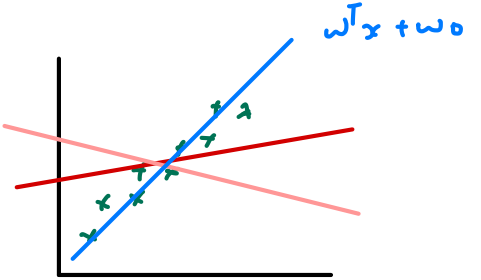
Linear Regression-03

Recap

$$\hat{y} = w^T x + w_0$$



MSE → loss func.

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)} \right)^2$$

$$\underset{w}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} \left(y^i - \hat{y}^i \right)^2$$

G.D:

1. Start randomly w
2. repeate {

$$w_j = w_j - \eta \cdot \frac{\partial \mathcal{L}}{\partial w_j}$$

}

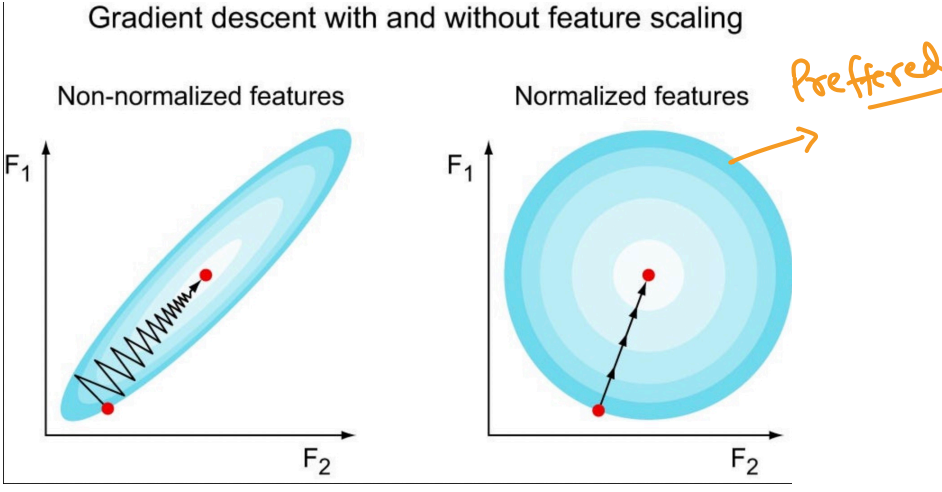
$$\frac{\partial \mathcal{L}}{\partial w_j} = \frac{2}{n} \sum_{i=1}^{n} (\hat{y}^i - y^i) \cdot x_j$$

Feature Scaling

1. importances

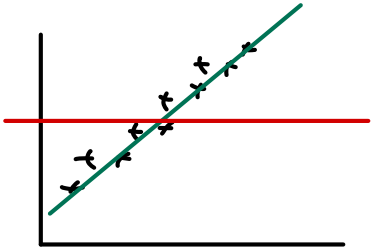
age [1 - 15] | odometer [1000 - 100000]

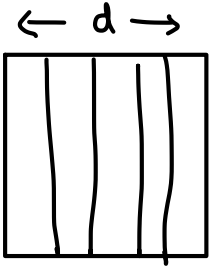
2. G.D converges faster



Adj R2 Score

$$R_2 \text{ score} = \left[1 - \frac{SS_{res}}{SS_{total}}\right]$$





$d+1$ →

Case I: feature is relevant ($R_2 \uparrow \uparrow$)

Case II: feature is irrelevant ($R_2 \uparrow$ or remain same)

wrong

$0.88 \rightarrow 0.89$

why ?

C.D → $w_j f_j$

$= 0$

2 features

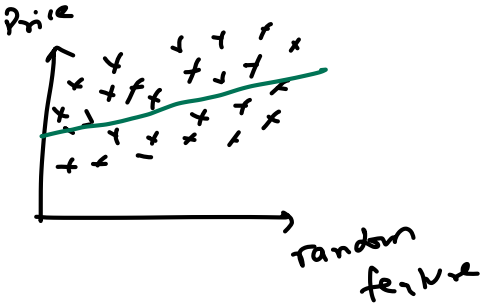
$w_1 x_1 + w_2 x_2$

↓

3 features

$w_1 x_1 + w_2 x_2 + w_3 x_3$

$= 0$



$w_j = 0.8$

R2 never decreases. [remain same or increases.]

Adj R_2 $\Rightarrow$ $1 - \left[\frac{(1-R_2)(n-1)}{n-d-1}\right]$

$n \rightarrow$ # points

$d \rightarrow$ # features

(1) if $d+1$ irrelevant

$\downarrow$

Adj R_2 $\downarrow$

$\frac{10}{5} = 2$

$\frac{10}{2.5} = 5$

(2) if $d+1$ relevant

$\downarrow$

Adj R_2 $\uparrow$

$d \uparrow$ = adj $R_2 \downarrow$

$d \uparrow$ (relevant) = $R_2 \uparrow\uparrow \Rightarrow$ adj $R_2 \uparrow\uparrow$

Net effect $\Rightarrow$ adj $R_2 \uparrow$

Theoretical

$R2 \rightarrow (-\infty, 1]$

$adj\ R2 \rightarrow (-\infty, 1]$

Real world

$R2 \rightarrow [0, 1]$

$adj\ R2 \rightarrow [0, 1]$

StatsModel

→ Statisticians

→ Statistical functionalities

OLS → "Ordinary Least Squares"

How is OLS different from sklearn Linear Regression?

OLS

Provides detailed statistical summary about

- goodness-of-fit
- p-values
- confidence intervals
- coefficients

To assess the quality of model, offers

- residual plots
- QQ plots
- influence statistics

Sklearn

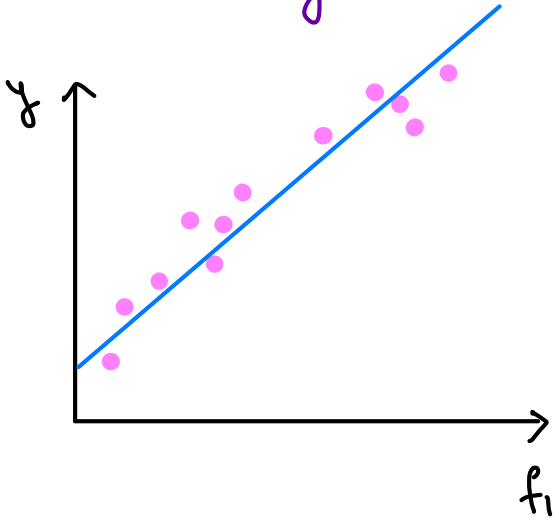
- Primary goal is to build a predictive model
- Less concerned about detailed statistical analysis and interpretation.
- Offers additional features and functionalities like :

➡ Feature scaling
➡ Regularization
➡ Cross validation
➡ Evaluation metrics

Assumptions of Linear Regression

- Assumption of Linearity
- No Multi-Colinearity
- Normality of Residuals $(y - \hat{y})$
- No Heteroskadasticity
- No Autocorrelation

① Linearity



"linear relationship"