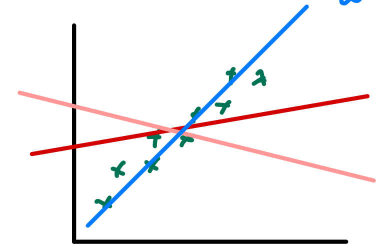
Linear Regression-03

Recap



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

MSE $\rightarrow$ loss func.

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

$$\underset{w, w_0}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} \left(y^i - \hat{y}^i \right)^2$$

G.D: ① Start randomly w

② repeate {

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

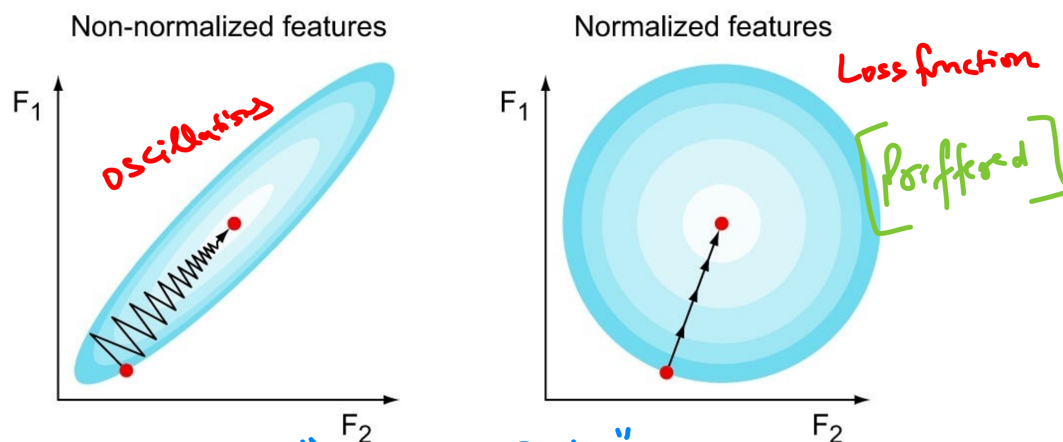
}

Feature Scaling

1. importances   age / odometer
   [1-15]   [1000-1L]

2. G.D Converges faster

Gradient descent with and without feature scaling



" Contour Plots "

Performance Metric

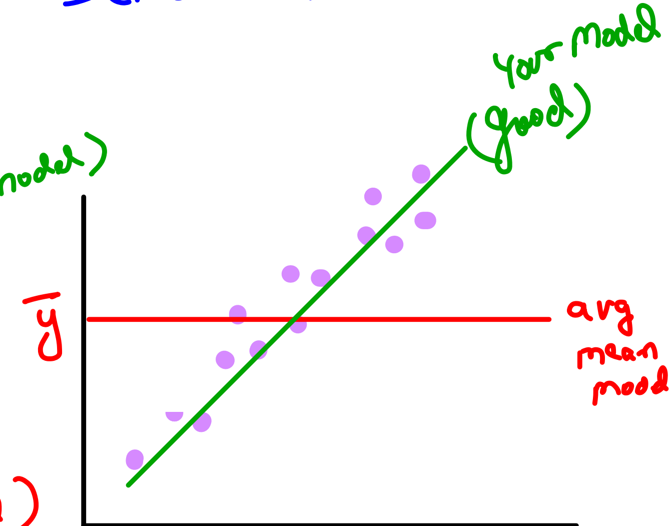
R2-Score / R-squared / Coeff. of Determination

~~accuracy~~ → Performance

(MSE of good model)

$$R2\text{-}Score = 1 - \frac{SS_{model}}{SS_{total}}$$

(MSE of avg model)



$\bar{y}$ = mean of (y-train)

$$R2\,Score = 1 - \frac{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\sum_{i=1}^{n} \left(y^{(i)} - \bar{y}\right)^2}$$

R2_Score

- $R2 = 1$

  $MSE_{model} = 0$

  [Perfect]

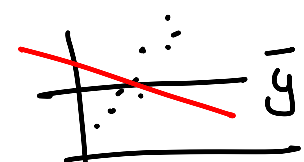
- $R2 = 0$

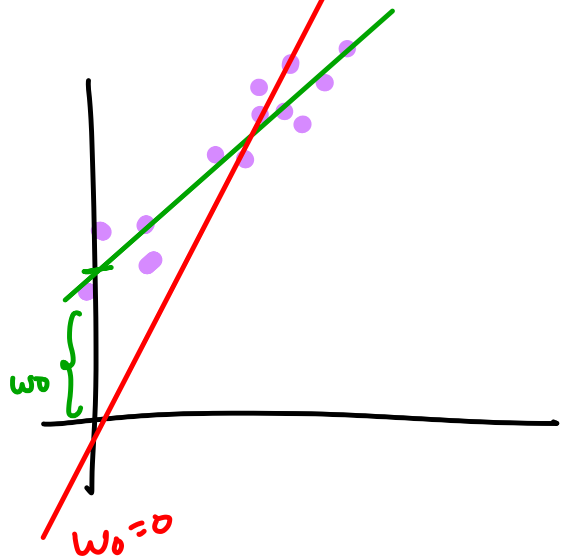
  Pred = avg

  Model = avg model

- $R2 = -ve$

  Worse than mean model



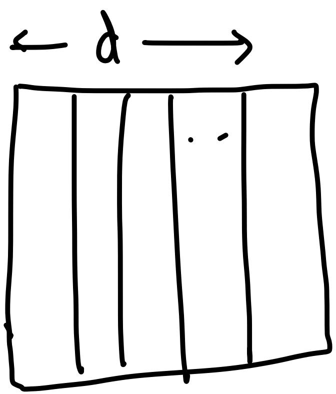


$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix}$$

$$w_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d$$

Adj R2 Score

$$R_2 = \left[1 - \frac{SS_{model}}{SS_{total}}\right]$$



$\xrightarrow{d+1}$

Case I : feature is relevant (R2 ↑ ↑) 0.88 → 0.91

Case II : feature is irrelevant (R2 ↑ or remain same)

0.88 → 0.89

$$\hat{y} = w_1x_1 + w_2x_2 + \cdots + w_dx_d + w_{d+1}x_{d+1}$$

$w_{d+1} \rightarrow \approx 0$

"gradients descent"

$$\text{Adj } R2 \Rightarrow 1 - \left[\frac{(1-R2)\cdot(n-1)}{n-d-1}\right]$$

n = # datapts
d = # feature
$R2$ = R2-score

① if $d+1$ irrelevant
Adj R2 ↓

② if $d+1$ relevant

Adj R2 ↑

$d\uparrow$ = adj R2 ↓
$d\uparrow$ (relevant) = R2 ↑↑↑ =

Net effect = adj R2 ↑

How does the adjusted R2 score differ from the regular R2 score?

3 users have participated

A	The adjusted R2 score accounts for the number of predictors in the model.	67%
B	The adjusted R2 score is always higher than the regular R2 score.	0%
C	The adjusted R2 score considers only the explanatory power of the model.	33%
D	The adjusted R2 score is not influenced by the sample size.	0%

End Quiz Now

In adjusted R-squared, what is the range of possible values?

58 users have participated

A	0 to 1	43%
B	-infinity to 1	33%
C	-1 to 1	14%
D	-infinity to infinity	10%

End Quiz Now

StatsModel

→ Statistician

→ statistical functionalities.

X, Y

$$\boxed{w_{Best} = (X^T X)^{-1} \cdot (X^T y)}$$

Normal eqn

closed form soln

OLS → ordinary least squares

How is OLS different from sklearn Linear Regression?

OLS — Statsmodel

Provides detailed statistical summary about

- goodness-of-fit
- p-values
- confidence intervals 95%
- 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 (m.imp)
  - ➡ Cross validation ✓
  - ➡ Evaluation metrics ✓

Why might the adjusted R2 score be considered more reliable than R2 when adding more predictors to a model?

43 users have participated

A	Because it always increases with more predictors.	12%
B	Because it penalizes the model for adding predictors that don't improve the model.	84%
C	Because it is easier to calculate.	2%
D	Because it always equals the R2 score..	2%

End Quiz Now

A regression model with 3 predictors has an R2 of 0.85. After adding a 4th predictor, the R2 increases to 0.86 but the adjusted R2 decreases. What can be inferred?

45 users have participated

A	The 4th predictor improved the model significantly.	7%
B	The adjusted R2 is incorrectly calculated.	2%
C	The 4th predictor did not add meaningful information to the model.	91%

End Quiz Now

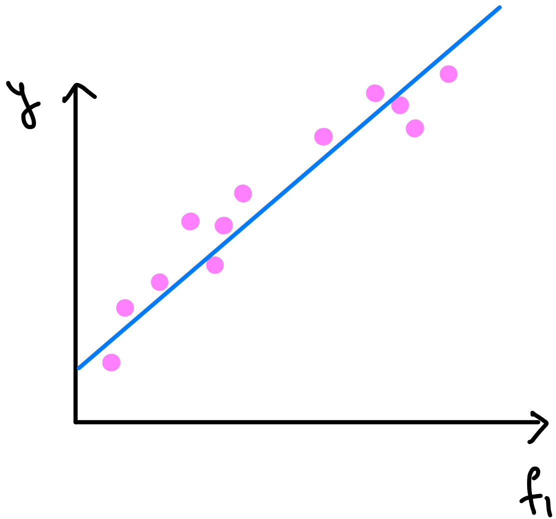
Assumptions of Linear Regression

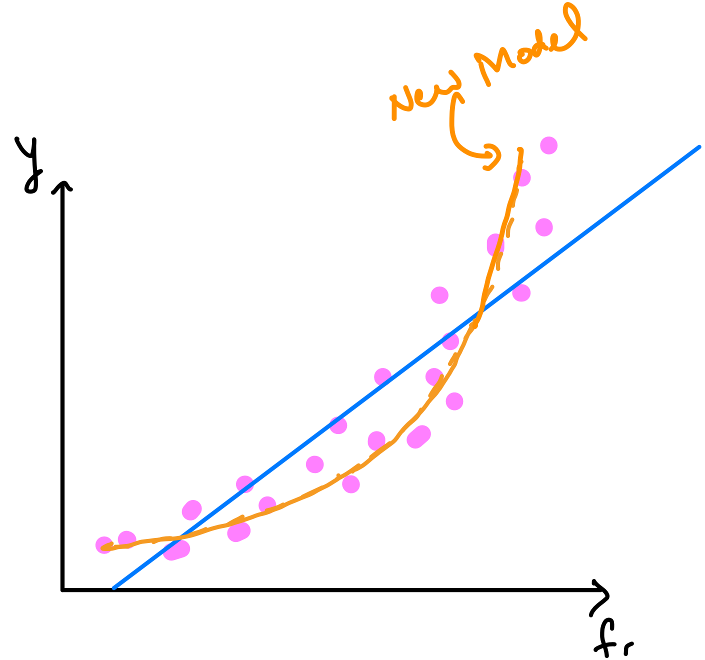
- → Assumption of Linearity
- → No Multi-Colinearity
- → Normality of Residuals $(y - \hat{y})$
- → No Heteroskadasticity
- → No Autocorrelation

$y - \bar{y}$

$16 - 14.5 \Rightarrow 1.5$

① Linearity –





"linear relationship"