Agenda:

1. Adjusted R^2 Score
2. Model Interpretation / Feature Imp.
3. Assumptions of linear Regression.
4. Cheat sheet of linear Regression.

1. Adjusted R^2 Score.

R^2 Score — value varies (0,1)

- $\to$ Sum-Squared Error (SSE) $\to$ Increases
- $\to$ residual Sum Squares (RSS) $\to$ decreases when features are added.

Bad reason whenever some new random feature.

$$\text{adj } R^2 \text{ Score} \rightarrow 1 - \frac{(1-R^2)(n-1)}{(n-p-1)}$$

$n \Rightarrow$ # examples
$p \Rightarrow$ # features
$R^2 \Rightarrow$ R-Squared Score

$n \Rightarrow 10$

R^2 Score $= 0.85$

Case 1: Add features (+) $\Rightarrow$ not useful features.

$$\text{adj } R^2 \rightarrow 1 - \underbrace{\frac{(1-0.85)(n-1)}{(n-p-1)}}_{\text{decrease}}$$

Case 2: Add features (+) $\Rightarrow$ v. imp. features.

$$R^2 \rightarrow 0.87 \quad (n=10)$$

$$\text{adj } R^2 \rightarrow 1 - \underbrace{\frac{(1-0.87)(n-1)}{(n-p-1)}}_{\text{increase}}$$

$\boxed{0.8}$ $\rightarrow$ adj $R^2 \downarrow$

Conclusion $\rightarrow$ adj $R^2 \uparrow$

$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots \ldots w_{10} x_{10}$$

(example)

$y = Pr(\text{Sales}) = w_0 + w_1 x_1$

weight $= \boxed{\sqrt{\text{Sales}^2 + R^2}}$

weight $= y_{\text{height}} + w_1 y_{\text{age}} \ldots \frac{w_{10}}{w_1}$

2. Interpretability & Feature Importance.

$$y = w_0 + w_1 x_1 + w_2 x_2 \ldots + w_{10} x_{10}$$

est. weights $2 < w_1 < w_2 \ldots w_3 \ldots \ldots w_{10}$

Case 1: $w_1 : \boxed{\text{small}}$

Reason $\rightarrow$ x_1 = millage $\qquad w_2 = 5$

$$y = w_0 + w_1 x_1 + \boxed{5} \text{ millage} + \ldots$$

$x_1 \rightarrow 2$ units

$x_2 \rightarrow \boxed{2}$ $\rightarrow$ units

$y(x_1, x_2) = 2 \text{ units}$

Case 2: $w_1 : \boxed{-ve}$

Reason $w_1 = -20$, $w_2 = \text{odometer reading}$.

low age $\qquad$ high x_1

$x_1 \uparrow$ (age) $\rightarrow$ low y (w.r.t)

Case 3: $w_1 : \boxed{0}$

x_1 = some random feature

no change on the $y(x_1)$!!!

$\boxed{0}$ $\qquad$ x_1 = some random feature
low w.r.t y
$w_1 = 0$
$y = w_0 + w_2 x_2 + w_3 x_3$
$w_1 \rightarrow 0$

Suppose $w_1 = 2$ & $\boxed{w_2 = 10}$

Suppose $w_1 = -2$ & $\boxed{w_2 = -5}$
$x_1 \uparrow$

target

Target mean differential Value + Bigger $|w_i|$

Bigger : $w_1 = 2$ $\qquad \boxed{w_2 = -9}$

Issues with the approach

Assume,

$$y = w_0 + w_1 x_1 + \ldots + (\text{random}) \underset{[0,1]}{x_2} + \ldots + (-w) \underset{[10,-10]}{x_3}$$

$[0,1]$

Solution $\rightarrow$ Scaling / Standardization / normalization

(1) Standardization $\rightarrow$ mean centering + variance scaling

$$x' = \frac{x - \mu}{\sigma} \qquad [-3, +3]$$

Pb in (1) $\rightarrow$ (2) Min-Max Scaler $\leftarrow$ Normally recommended

$$\frac{x - \min(x)}{\max(x) - \min(x)}$$

weights $\rightarrow$ compensate.

$$y = w_0 + w_1 x_1 + \ldots \underbrace{(-w)}_{[-10,-10]} x_3 + \ldots + (0.1) x_2$$

Point features.

$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots w_{10} x_{10}$

$x_1 \rightarrow$ some random feature
$x_2 \rightarrow$ some random feature
$x_3 \rightarrow$ some random feature
$x_4 \rightarrow$ some random feature

$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots w_{10} x_{10}$

x_1 = some highly correlated
x_2 = x_1 + Gaussian noise
x_3 = $x_1 + x_2$

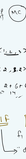
$w_1 \rightarrow 0$

3. Assumptions of Linear Regression.

Pick Sample $\rightarrow$ Regression & Interpretation (Inference)

(Prediction)

(1) Assumption of linearity.



$y = f_1(x)$ long

$y = w_0 + w_1 x + w_2 x^2$

$y = w_0 + w_1 x_1 + w_2 x_2$

(2) No Multicollinearity (Image Factory)

Linearity

x_1 & x_2

s.t. $x_1 = w_1 x_2 + \beta$

Independency are better or Strongly other factor with very linear dependencies.

$y = 2x_1 - 3x_2 + \epsilon$

$+ \beta$

$\begin{bmatrix} 2 \\ -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Multicollinearity

$\rightarrow$ Linearity arises multiple factors.

$\boxed{x_1}$ $\boxed{x_2}$ $\boxed{x_3}$

$\boxed{x_4}$ $\quad w_1 x_1 + w_2 x_2 + w_3 x_3 + \beta$

Perfect Correlation

Why not x_4 is a feature?

Reason $\rightarrow$ $x_1 = w_1 x_2 + \beta$

Ex $\langle x_1, x_2, x_3 \rangle$ $\qquad [0,1,2]$

$y_1 = \langle 0, y_2, y_3 \rangle$

$y_2 = 1 \cdot x_1 + \boxed{0} x_2 + y_3$

$\Rightarrow 1 \cdot x_1 + 2 \cdot (1 \cdot x_1) + 1 \cdot y_3$

$\Rightarrow 1 \cdot x_1 + 2 \cdot x_1 + y_3$

$\Rightarrow 3 \cdot x_1 + y_3$

$\Rightarrow 3 \cdot x_1 + 0 \cdot y_2 + y_3$

$y_2 = \langle 1, 1, 1, 3 \rangle$

$y_3 = \langle 1, 2, 1, 3 \rangle$

Reason x_2 by x_1 (Linearly) $\rightarrow$ $\boxed{y_3 = f(x)}$

$\rightarrow$ $\boxed{\langle x_1, 0, y_3 \rangle}$

$y_1 = \langle 1, 1, 1, 3 \rangle$

$y_2 = \langle 1, 2, 1, 3 \rangle$

$\rightarrow$ Linearly $\rightarrow$ x_1 $\rightarrow$ x_2

$y_3 = \langle 1, 1, 1, 3 \rangle$

$\rightarrow$ Linearly $\rightarrow$ x_2 $\rightarrow$ x_3

$\rightarrow$ $\boxed{\langle x_1, 0, y_3 \rangle}$

$\rightarrow$ x_1 $\rightarrow$ x_3

$\rightarrow$ $\boxed{\langle x_1, 0, y_3 \rangle}$

$\rightarrow$ x_1 $\rightarrow$ x_2

$\rightarrow$ x_2 $\rightarrow$ x_3

$\rightarrow$ x_1 $\rightarrow$ $x_3