Linear-Regression-2

Monday, 10 October 2022 7:41 PM

Agenda :-

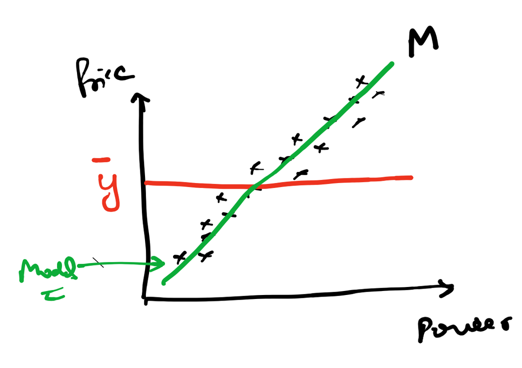
1. Evaluation Metric (R2 Score)
2. Sklearn implementation
3. Multiple Linear Regression.
4. Scratch Code
5. Evaluation Metric (Adj R2 Score)
6. feature Importance / Interpretation. [skip]

Evaluation Metric (R2 Score)

↳ Performance of the Model.

'new' car → $\boxed{M}$ → $\hat{y}$

M_1 M_2



x	y

↳ $\bar{y}$

x_{new} 'xq' → Dumb Model. $\boxed{M_{Dumb}}$ → $\bar{y}$ (mean prediction) 8.752

R2 Score (R^2) → $1 - \dfrac{SS_{model}}{SS_{Total}}$

SS_{model}: MSE Errors of linear Reg model

SS_{Total}: ↳ MSE errors of dumb

$$SS_{model} = \frac{1}{n}\sum_{i=1}^{n}\left[y^{(i)} - \hat{y}^{(i)}\right]^2$$

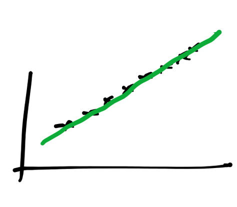
↳ Predictions from Linear Reg. Model

$$SS_{total} = \frac{1}{n}\sum_{i=1}^{n}\left[y^{(i)} - \bar{y}\right]^2$$

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

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

Range of R2



1. Best value of R^2? ⇒ 1

$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$ = when $\boxed{SS_{res} = 0}$

2. when $R^2 = 0$?

$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$ ⇒ 0

$SS_{res} = SS_{total}$ (dumb a Model)

3. R2 as negative?

→ YES

$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$ >> 1

$SS_{res} > SS_{total}$

(−∞ 0 1]

Q→ What is a good R^2?

Case I: ~~Cat~~ Car Price Prediction. 0.92 = good.

Case II: Cancer Severity Prediction. $R^2 \Rightarrow 0.98 - 0.99 \ldots$

Multiple Linear Regression (Multi-Variate)

$$\hat{y} = w_0 + w_1 x_1$$

d input variable.

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots + w_d x_d$$

$w = \langle w_1, w_2 \ldots w_d \rangle$

$x = \langle x_1, x_2, x_3 \ldots x_d \rangle$

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

$x_0 = 1$, $x_1\ x_2\ x_3 \ldots x_d$, y

$\hat{y} = \boxed{w_0} x_0 + w_1 x_1 + \ldots w_d x_d$

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

$w = \langle w_0, w_1 \ldots w_d \rangle$

$x = \langle x_0, x_1, \ldots x_d \rangle$

d-features

$W \to (d, 1)$

$X \to (n, d)$

↳ n

Predict $(X, w) \to n$ predictions.

$W_{(d,1)}$ $X_{(n,d)}$ $\boxed{(n,1)}$

$X \cdot W$

$(n,d)\ (d,1)$

$\hat{Y} \Rightarrow (n,1)$

Predict.

$$\boxed{\hat{Y} = np.dot(X, w)}$$

@ 10:39.

$X \cdot W$: $(n,d)\ (d,1) \to (n,1)$ & $W^T X^T$: $(1,d) \cdot (d,n) \to (1,n)$

$\boxed{X \cdot W}$ (n-sample, n-features) → (n-features, 1)

$w^T x$ → for 1 example.

$(1,d)\ (d,1) \to (1,1)$

$x^{(i)} \Rightarrow \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}_{(d,1)}$ $w = \begin{bmatrix} \ \end{bmatrix}_{(d,1)}$

Error function:

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

y $(n,1)$ − $\hat{y}$ $(n,1)$ ⇒ np.mean($[\ ]$ ** 2) $(n,1)$ ⇒

⇓

real value.

3.82

(n,d) $(d,1)$

$X \to (n,d)$ $y^{(n,1)}$ Price

$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}_{(d,1)}$

Gradient.

$$\frac{\partial L}{\partial w_0} \Rightarrow \frac{2}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial L}{\partial w_1} \Rightarrow \frac{2}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right) \cdot x^{(i)}$$

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

$$\boxed{\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2 + \ldots w_d x_d}$$

$\frac{\partial \mathcal{L}}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2} \ldots \boxed{\frac{\partial L}{\partial w_j}} \ldots \frac{\partial L}{\partial w_d}$

$$\frac{\partial L}{\partial w_j} = \frac{2}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right) \cdot \frac{d}{dw_j}\left[y^{(i)} - \hat{y}^{(i)}\right]$$

$$\frac{d}{dw_j}\left[\cancel{y^{(i)}} - \cancel{w_0 x_0} - \cancel{w_1 x_1} \boxed{- w_j x_j} \ldots - \cancel{w_d x_d}\right]$$

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

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

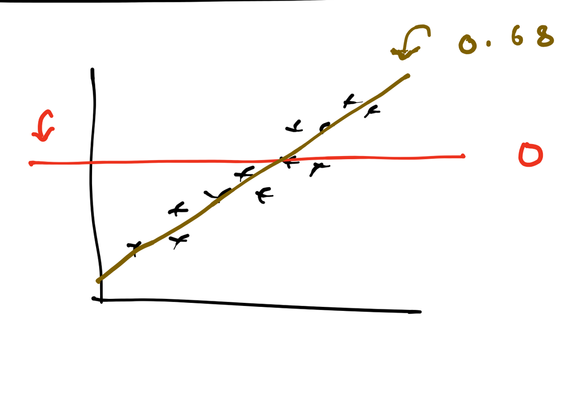
$Y_{(n,1)}$ $\hat{Y}_{(n,1)}$ $X_{(n,d)}$ W ↳ $(d,1)$

$\underbrace{(y\_hat - y)}_{(n,1)}$ $X_{(n,d)}$

$X^T_{(d,n)} \cdot (\hat{Y} - Y)_{(n,1)}$

$(d,1)$

Doubts



OHE ⇕ target Encoding → Nominal ↳ "Pincode"

Label Encoding → ordinal Data ↳ Grades, $\boxed{A, A^+, B, B^-, C, D}$