Linear-Regression-2

Wednesday, 3 August 2022 8:52 PM

Agenda :

0. Sklearn implementation
1. Multi-variate Linear Regression
2. Evaluation Metric → adj R^2
3. feature Imp. /weights
4. Assumptions of Linear Regression

$X \rightarrow$ (n-samples , n-features)

$X \rightarrow$ (19820 , 1)

$Y \rightarrow$ 1D

that's fine

(19820 ,)

Multivariate Lin. Reg. (Multiple Linear Regression)

$f_1 \; f_2 \; f_3 \; f_4 \dots f_d$

$\hat{y}$

hypo $\Rightarrow f(x) = w_1x_1 + w_2x_2 + \dots + w_dx_d + w_0x_0$ → 1

$\Rightarrow \boxed{w^Tx + w_0}$

$\Rightarrow \sum_{i=0}^{d} w_ix_i$

x_0 ← d → ; n ; 1 ; n

X ↳ (n, d+1)   Y ↳ (n,1)

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

$\hat{Y} \Rightarrow X.W$ (n, d+1) (d+1, 1)

$\hat{Y}$ → (n,1)

np.dot (X, w)

$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} \cdot \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

$$\begin{bmatrix} x_0^{(1)}w_0 + x_1^{(1)}w_1 + x_2^{(1)}w_2 + \dots x_d^{(1)}w_d \\ \\ \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(n)} \end{bmatrix}_{(n\times 1)}$$

Error function

Loss / MSE $\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2$

$(y-\hat{y})**2$

Y [ ] − $\hat{Y}$ [ ] ⇒ [ ].mean()

↓ Single Real value Loss value

Gradients.

w_0 & w_1

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

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

$f(x) = w_0x_0 + w_1x_1 + \dots w_jx_j + \dots w_dx_d$

$\frac{\partial L}{\partial w_0}$ $\frac{\partial L}{\partial w_1}$ $\frac{\partial L}{\partial w_2}$ $\boxed{\frac{\partial L}{\partial w_j}}$ . . . $\frac{\partial L}{\partial w_d}$

$$\frac{\partial L^{(i)}}{\partial w_j} \Rightarrow 2\left[y - \hat{y}\right] \times \frac{\partial\left[\cancel{y} - \overbrace{\cancel{w_0x_0} - \cancel{w_1x_1} - \cancel{w_2x_2} \dots \boxed{w_jx_j} \dots \cancel{w_dx_d}}^{\hat{y}}\right]}{\partial w_j}$$

$\frac{\partial L^{(i)}}{\partial w_j} \Rightarrow 2\left[y - \hat{y}\right] \times (-x_j)$

$\frac{\partial L^{(i)}}{\partial w_j} \Rightarrow 2\left[\hat{y} - y\right].x_j$

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

$j=0$

$\frac{\partial L}{\partial w_0} = \frac{2}{n}\sum_{j=1}^{n}\left[\hat{y}^{(i)} - y^{(i)}\right].x_0^{(i)}$

$\boxed{X}$ ↳ (n, d+1) , $\boxed{Y}$ ↳ (n,1) $\boxed{\text{weights}}$ ↳ (d+1, 1)

grad ↳ (d+1, 1)

$$\underbrace{X^T}_{(d+1,\,n)} . \left[\underbrace{\hat{Y} - Y}_{(n,1)}\right] \rightarrow (d+1, 1)$$

← d+1 →

$$X = \begin{bmatrix} — x^{(1)} — \\ — x^{(2)} — \\ \vdots \\ — x^{(n)} — \end{bmatrix}_{(n,\,d+1)}$$

$$X^T \Rightarrow \begin{bmatrix} x^{(1)} & x^{(2)} & \dots & x^{(n)} \end{bmatrix}_{(d+1,\,n)} \cdot \begin{bmatrix} \overbrace{w_0x_0^{(1)} + w_1x_1^{(1)} + \dots + w_dx_d^{(1)}}^{\hat{y}} - y^{(1)} \\ w_0x_0^{(2)} + w_1x_1^{(2)} + \dots + w_dx_d^{(2)} - y^{(2)} \\ \vdots \\ w_0x_0^{(n)} + w_1x_1^{(n)} + \dots + w_dx_d^{(n)} - y^{(n)} \end{bmatrix}_{(n,1)}$$

$$\begin{bmatrix} \hat{y}^{(1)} - y^{(1)} \\ \hat{y}^{(2)} - y^{(2)} \\ \hat{y}^{(3)} - y^{(3)} \\ \vdots \\ \hat{y}^{(n)} - y^{(n)} \end{bmatrix}$$

$$\boxed{\frac{\partial L}{\partial w_0}} \Rightarrow \left(\hat{y}^{(1)} - y^{(1)}\right) \times x_0^{(1)} + \left(\hat{y}^{(2)} - y^{(2)}\right) \times x_0^{(2)} + \dots + \left(\hat{y}^{(n)} - y^{(n)}\right) \times x_0^{(n)}$$

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

Adj. R2 Score

+1 feature

- → if feature is relevant : R^2 increase
- → if feature is totally irrelevant : R^2 incr. or remain same.

$\boxed{f_j}$ → temp. of city (delhi) → irrelevant

$w_j \approx 0$

→ "spurious relation/association"

↓

R2 Score

Price ⇒ −10 × odometer   100   1000

Price = $\boxed{0.9}$ × #icecream #ices

adjusted R2 Score $\Rightarrow 1 - \frac{(1-R^2)(n-1)}{(n-d-1)}$

Case I : d increases : but not significant

Hence adj R2 ↓ (reduce)

Case II : d increases : they are good features → R^2 will improve a lot.

Hence adj R2 ↑ (incr)

Feature Importance

$\hat{y} = w_0x_0 + w_1x_1 + \dots + w_dx_d$

⇒ weights $\langle w_0, w_1, \dots w_d \rangle$

Case I : $\boxed{w_j \Rightarrow +ve}$

Assume f_2 = Mileage & $w_2 = 2$

$\hat{y} = w_0x_0 + w_1x_1 + \boxed{2 \times \text{mileage}} + \dots + w_dx_d$

m=15   m=16

Case II : $\boxed{w_j = 0}$

f_j has no effect/impact on the price.

Case III : $\boxed{w_j = -ve}$

f_j : odometer reading ; $w_j = -10$

Suppose : $w_1 = 2$ & $w_2 = 10$

$\hat{y} = w_0 + 2f_1 + 10f_2 + \dots + w_dx_d$

f2 is more imp.

Suppose $w_1 = -5$ $w_2 = -10$

$\hat{y} = w_0 - 5f_1 - 10f_2 + \dots + w_dx_d$

f2 is more imp.

Suppose $w_1 = -8$ → f_1 $w_2 = +4$ → f_2

f1 is more imp.

Biggest influenced feature → $|w_j|$

Issue

$\hat{y} = w_0 + \dots + (-100{,}000) \times \text{age} + \dots + (-10) \times \text{odometer}$

age → [1, 15]

Sol^n → Col^m standardization → mean centering + var scaling

mean centering → 0   var scaling → 1

Data → $\boxed{\text{Col. Standardise}}$ → $\boxed{\text{Model fitting}}$ →

w_1, w_2 , (−1.8) age , ... (−2.3) odometer.

→ odometer has more importance

feature importance.