Polynomial Regression

Wednesday, 15 December 2021 7:53 PM

$h_\theta(x) = \boxed{\theta^T x} \to X \cdot \theta$

$X = \begin{bmatrix} \vdots \end{bmatrix}_{(m, n+1)}$

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}_{(n+1, 1)}$

$\boxed{m, 1}$

$\theta^T x$

$\boxed{\hat{y} = X \cdot \theta}$ ← np.dot (X, theta)

$(m,n) \; (n,1) \to (m,1)$

$\frac{X - \mu}{\sigma}$ / why ?

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

$(5 - 6)$

$\xleftarrow{\quad n \quad}$

$x_1 \quad x_2 \quad \ldots \quad x_n$

$[\;]_{(m,n)} - [\;]_{(m,n)} \; (1,n) \Rightarrow [\;]_{(m,n)}$

std $\boxed{(n,1)}$ m,n

Error : $\frac{1}{2m} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2$

$\underset{(m,1)}{\overset{y}{[\;]}} - \underset{(m,1)}{\overset{\hat{y}}{[\;]}} \Rightarrow \text{np.sum}\left([\;]_{(m,1)} ** 2 \right) \Rightarrow \frac{\text{real value}}{2 \times m}$

Gradient

$$\frac{\partial L}{\partial \theta_j} = \boxed{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot x_j^{i}}$$

(m,n)

$\text{grad} = \left\{ \begin{bmatrix} \frac{\partial L}{\partial \theta_0} \\ \frac{\partial L}{\partial \theta_1} \end{bmatrix}_{(n,1)} \right\} \Rightarrow \underset{(m,1)}{\overset{\hat{y}}{[\;]}} - \underset{(m,1)}{\overset{y}{[\;]}} \Rightarrow \overset{\hat{y} - y}{[\;]}_{(m,1)}$

(m,1)

$\boxed{X}_{(m,n)}$

(n,1)

$\boxed{X^T \cdot (\hat{y} - y)}$

$(n,m) \times (m,1) \to (n,1)$

theta = theta − lr. grad

$[\;]_{(n,1)} = [\;]_{(n,1)} - 0.1 [\;]_{(n,1)}$

$\text{grad} = \boxed{X^T (\hat{y} - y)}$

$\begin{bmatrix} \frac{\partial J}{\partial \theta_0} \\ \vdots \\ \frac{\partial J}{\partial \theta_n} \end{bmatrix} \Rightarrow \begin{bmatrix} x^{(1)} & x^{(2)} & \ldots & x^{(i)} & \ldots & x^{(m)} \end{bmatrix}_{(n,m)} \cdot \begin{bmatrix} \theta_0 x_0^{(1)} + \theta_1 x_1^{(1)} + \ldots + \theta_n x_n^{(1)} \\ \theta_0 x_0^{(2)} + \theta_1 x_2^{(2)} + \ldots + \theta_n x_n^{(2)} \end{bmatrix}_{(m,1)} - \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}_{(m,1)}$

$x^{(1)}$

$\left(\left[\theta_0 x_0^{(1)} + \theta_1 x_1^{(1)} + \theta_2 x_2^{(1)} + \ldots + \theta_n x_n^{(1)} \right] - y^{(1)} \right)$

$\left\{ \begin{bmatrix} x_0^{(1)} \to 1 \\ x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix} \right\} \quad \begin{bmatrix} x_0^{(1)} & x_0^{(2)} \\ x_1^{(1)} & \\ x_2^{(1)} & \\ \vdots & \\ x_n^{(1)} & \end{bmatrix} \quad] \times [$

accuracy = $\frac{\text{\# times you predicted correct}}{m}$

① Suzuki Alto $\boxed{1L}$: $\boxed{90K}$? $\boxed{99K}$

Actual Predicted

$\boxed{115K}$?

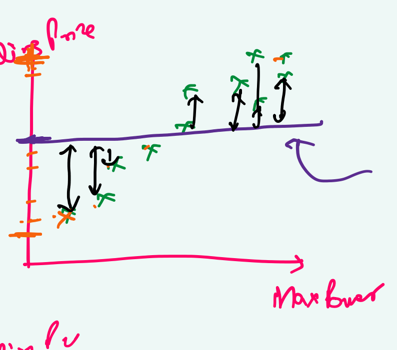
② Merc 64L : 63.9L

actual Predicted

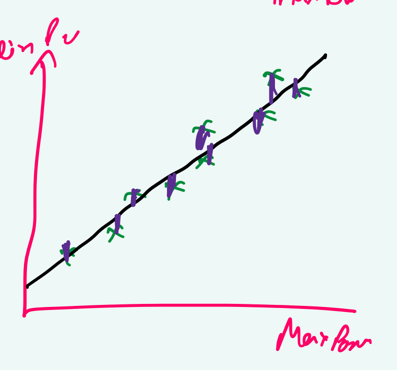
Closeness → MSE → 0.91, 105.23

Mean of all cars selling Price

$\bar{y}$ → Dumb Model



Linear Regression Model (M)



$$\boxed{R2 = \frac{MSE_{mean} - MSE_{regression}}{MSE_{mean}}}$$

R2 Score, R Squared,

Coefficient of Determination.

$$R2 = 1 - \frac{MSE_{reg}}{MSE_{mean}}$$

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

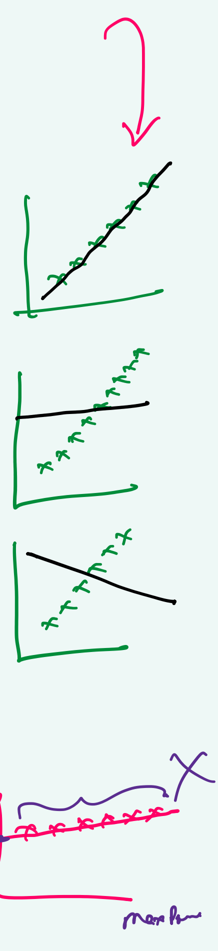
R2 Score

→ +1 (Perfect Score)

→ 0 (dumb)

→ −1 (even worse)

[0 ≤ R2 Score ≤ 1] → typical setting



$\boxed{-1 \leq R2 \leq 1}$ theoretical

Metrics

→ MSE

→ RMSE ⇒ $\sqrt{MSE}$

→ MAE

forget → {MAPE}

→ R2 Score (disadvantage)

→ $\boxed{\text{Adj. R2 Score}}$ (H.W)

$h(x) = \theta_0 + \theta_1 x$

$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2$?

$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$

Polyfeature(3)

$[x^0, x^1, x^2, x^3]$

$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \ - - - + \theta_{31} x^{31}$

$\boxed{\text{Train Data}} \to \boxed{\text{ML Algo}} \to \boxed{\text{hypothesis}} \to \text{Prediction}$

New Data → hypothesis

$\boxed{\text{Generalization}} \to$ 19K

New Cars

test data $\boxed{[\quad \text{Sample} \quad]}$ → hypothesis

test (actual selling Price)

Train | Test

$\boxed{\longleftarrow \; 80\% \; \longrightarrow \mid \longleftarrow 20\% \longrightarrow}$

Random Splitting 19980