University of Regina

ENEL/ENSE 865: Applied Machine Learning

Professor: Dr. Abdul Bais

Instructional Designer/TA: Dr. Muhammad Hamza Asad

Submitted by:

Name: Nandish Bakulkumar Bhatt
Student ID: 200441204

* Closed Form Solution:-

In program in function, simplin reg (input_feature, output): we used one eq^n to find slope and intercept. From here, the eq^n came, derivation is as below.

→ Now, we know that;

$$RSS(w_0, w_1) = \sum_{i=1}^{N} \left[y_{i\,(actual)} - y_{i\,(predicted)} \right]^2$$

$$= \sum_{i=1}^{N} \left[y_i - (w_0 + w_1 x_i) \right]^2$$

→ Now, doing partial derivate -

$$\frac{\partial RSS}{\partial w_0}(w_0, w_1) = \sum_{i=1}^{N} 2\left[y_i - (w_0 + w_1 x_i) \right](-1)$$

$$= (-2)\left[\sum_{i=1}^{N} y_i - N w_0 - w_1 \sum_{i=1}^{N} x_i \right]$$

$$\Rightarrow \frac{\partial RSS(w_0, w_1)}{\partial w_1} = 2 \sum_{i=1}^{N} \left[y_i - (w_0 + w_1 x_i) \right](-x_i)$$

$$= (-2)\left[\sum_{i=1}^{N} (x_i y_i) - \sum_{i=1}^{N} (w_0 x_i + w_1 x_i^2) \right]$$

∴ Gradient vector,

$$\nabla RSS(\hat{w}_0, \hat{w}_1) = \begin{bmatrix} (-2) \sum_{i=1}^{N} \left[y_i - (w_0 + w_1 x_i) \right] \\ (-2) \sum_{i=1}^{N} \left[y_i - (w_0 + w_1 x_i) \right] x_i \end{bmatrix}$$

$$= \begin{bmatrix} (-2)\left[\sum_{i=1}^{N} y_i - Nw_0 - w_1 \sum_{i=1}^{N} x_i\right] \\ (-2)\left[\sum_{i=1}^{N} x_i y_i - w_0 \sum_{i=1}^{N} x_i - w_1 \sum_{i=1}^{N} x_i^2\right] \end{bmatrix}$$

$\rightarrow$ Let, $s = \sum_{i=1}^{N}$

$$\therefore \nabla RSS(\hat{w}_0, \hat{w}_1) = \begin{bmatrix} (-2)(sy - Nw_0 - w_1 sx) \\ (-2)(sxy - w_0 sx - w_1 sxx) \end{bmatrix}$$

$\rightarrow$ To solve for w_0 and w_1, equate gradient to zero.

$$\therefore \nabla RSS(\hat{w}_0, \hat{w}_1) = 0$$

$$\therefore \begin{bmatrix} (-2)[sy - N\hat{w}_0 - \hat{w}_1 sx] \\ (-2)[sxy - \hat{w}_0 sx - \hat{w}_1 sxx] \end{bmatrix} = 0$$

$$\therefore N\hat{w}_0 + \hat{w}_1 sx = sy \quad \text{———} \quad (1)$$

$$\hat{w}_0 sx + \hat{w}_1 sxx = sxy \quad \text{———} \quad (2)$$

By eq^n (1),

$$\therefore \boxed{\hat{w}_0 = \frac{sy - \hat{w}_1 sx}{N}} \quad \text{———} \quad (3)$$

Put eq^n (3) into eq^n (2),

we get–

$$\left[\frac{sy - \hat{w}_1 sx}{N}\right] sx + \hat{w}_1 sxx = sxy$$

$$\hat{w}_1 \left[\frac{N sxx - sx * sx}{\cancel{N}} \right] = \frac{N sxy - sx * sy}{\cancel{N}}$$

$$\therefore \boxed{\hat{w}_1 = \frac{N sxy - sx * sy}{N sxx - sx * sx}} \quad \text{———} \quad (4)$$

→ Thus,

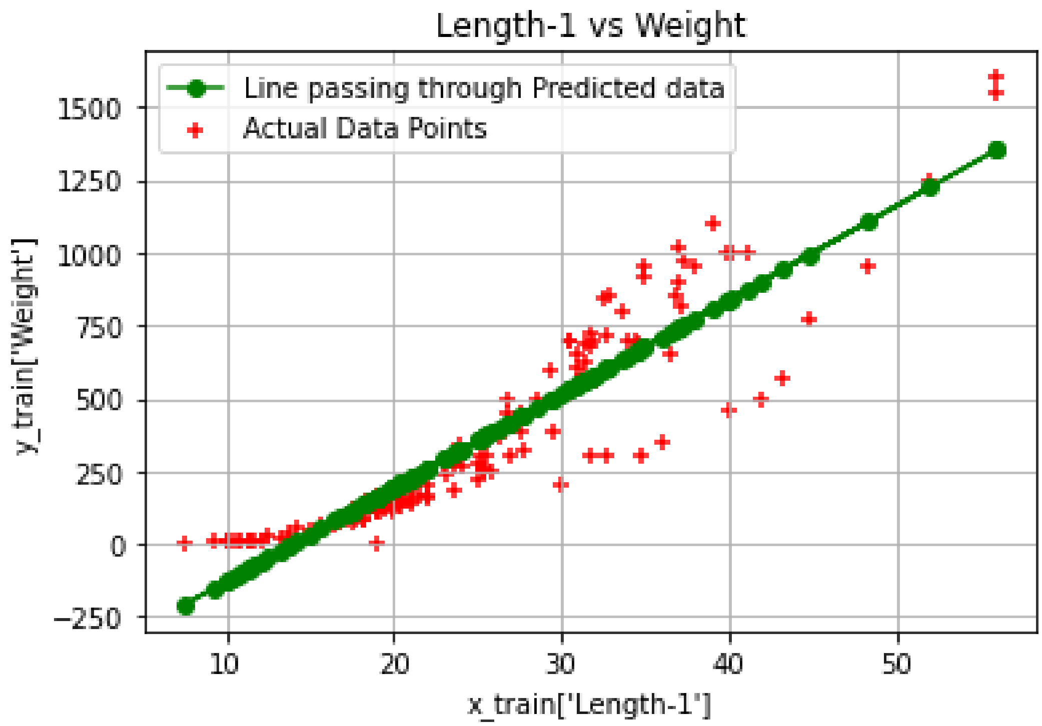
Solution Vector, $\hat{w} = \begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix} = \begin{bmatrix} \dfrac{sy - \hat{w}_1 sx}{N} \\ \\ \dfrac{N sxy - sx * sy}{N sxx - sx * sx} \end{bmatrix}$

→ Now, in program,

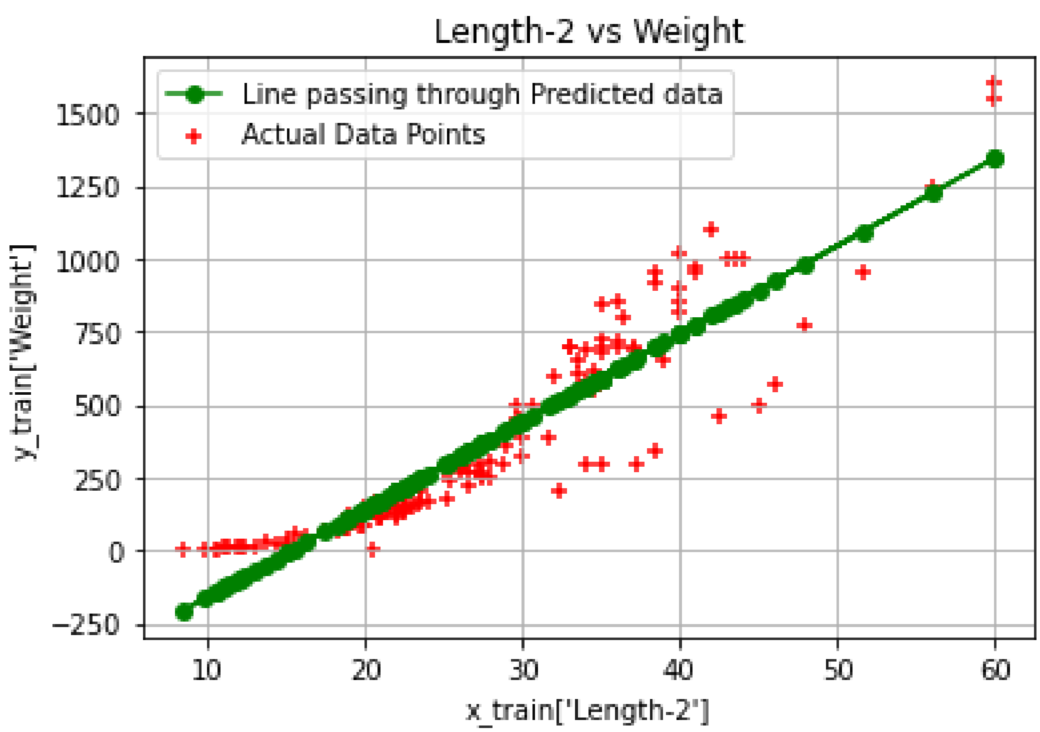
$$y_i = \underbrace{w_0}_{\text{Intercept}} + \underbrace{w_1}_{\text{Slope}} x_i$$

→ The values of intercept and slope can be found from the above solution vector.

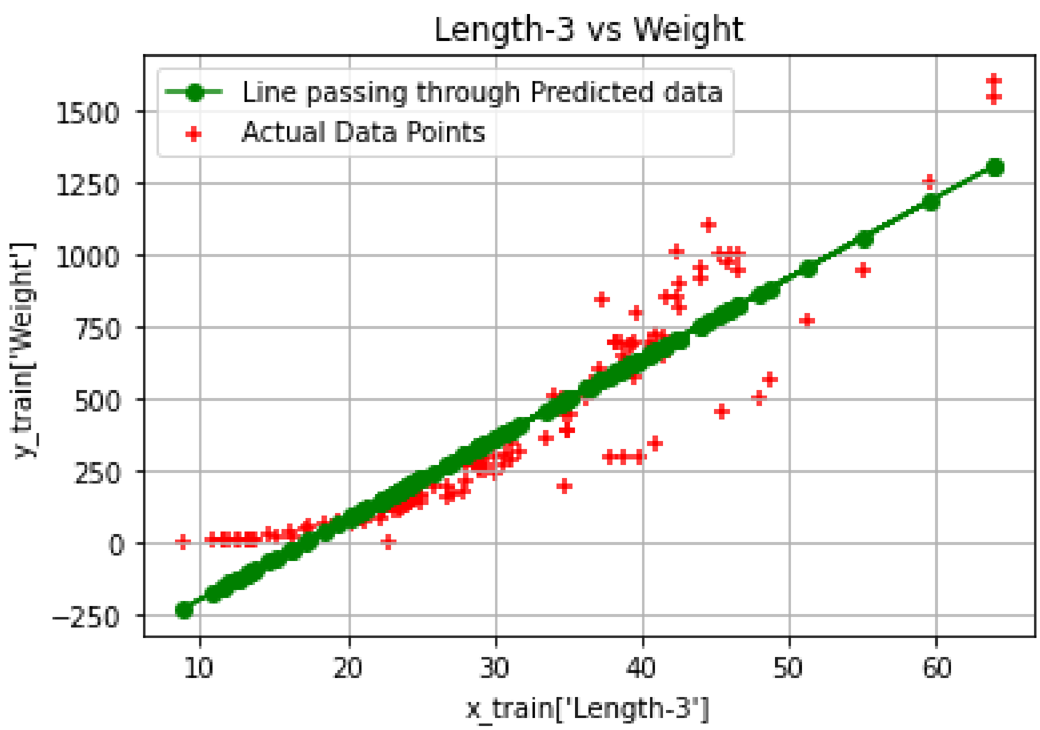
Results:

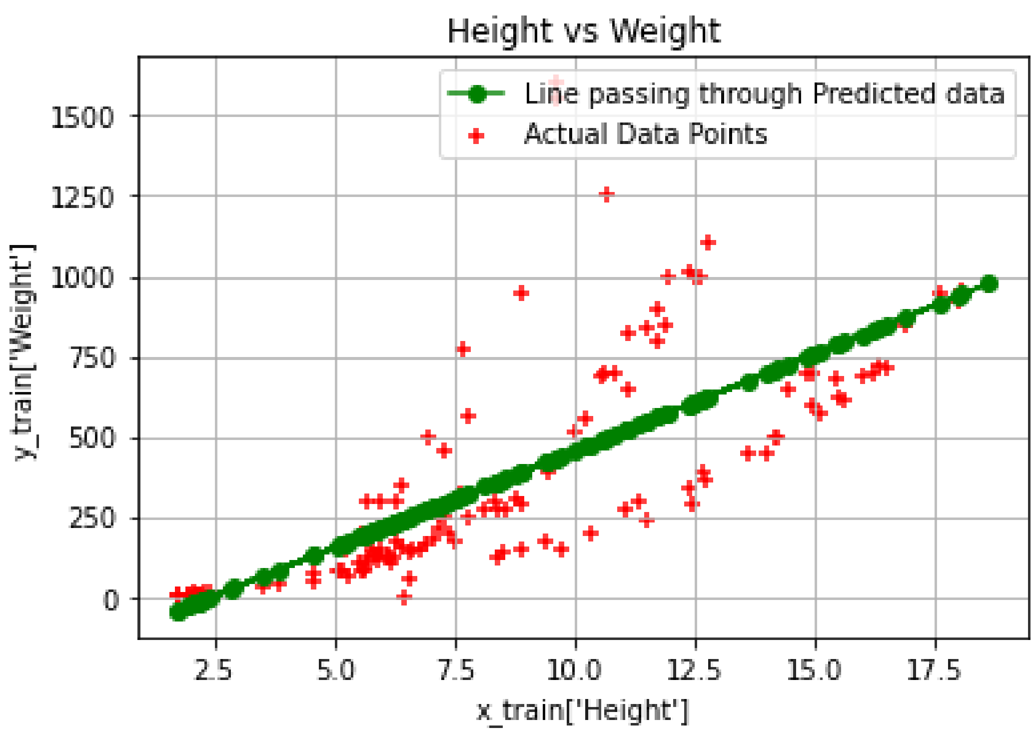


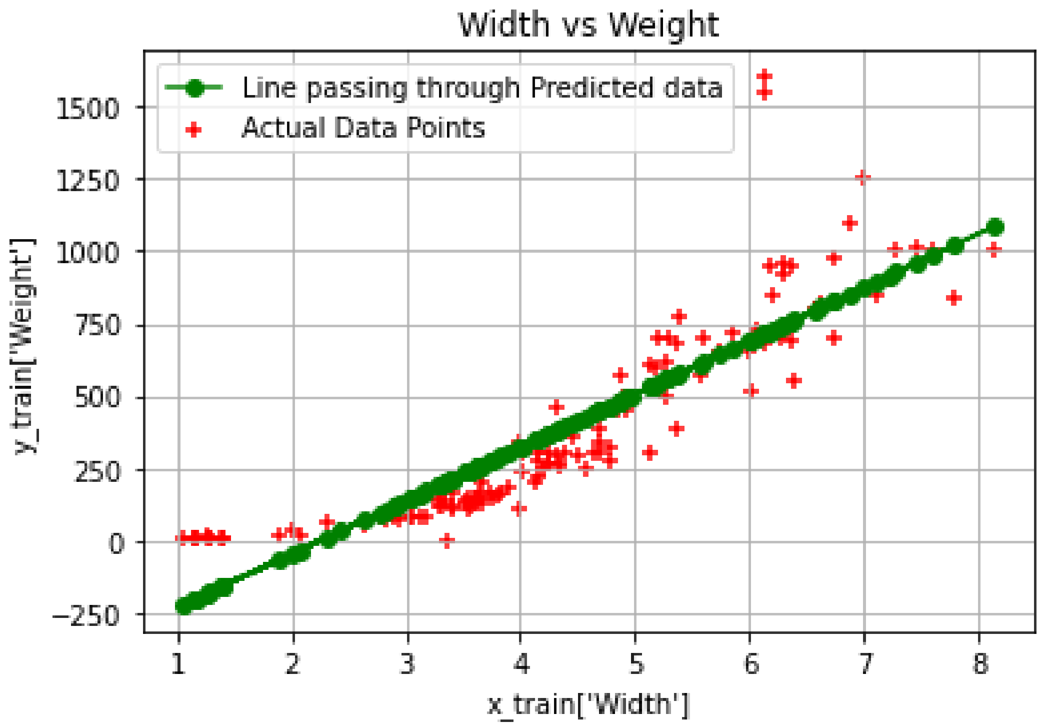
Length-1 vs Weight
Line passing through Predicted data
Actual Data Points
1500
1250
1000
750
500
250
0
−250
10
20
30
40
50
y_train['Weight']
x_train['Length-1']

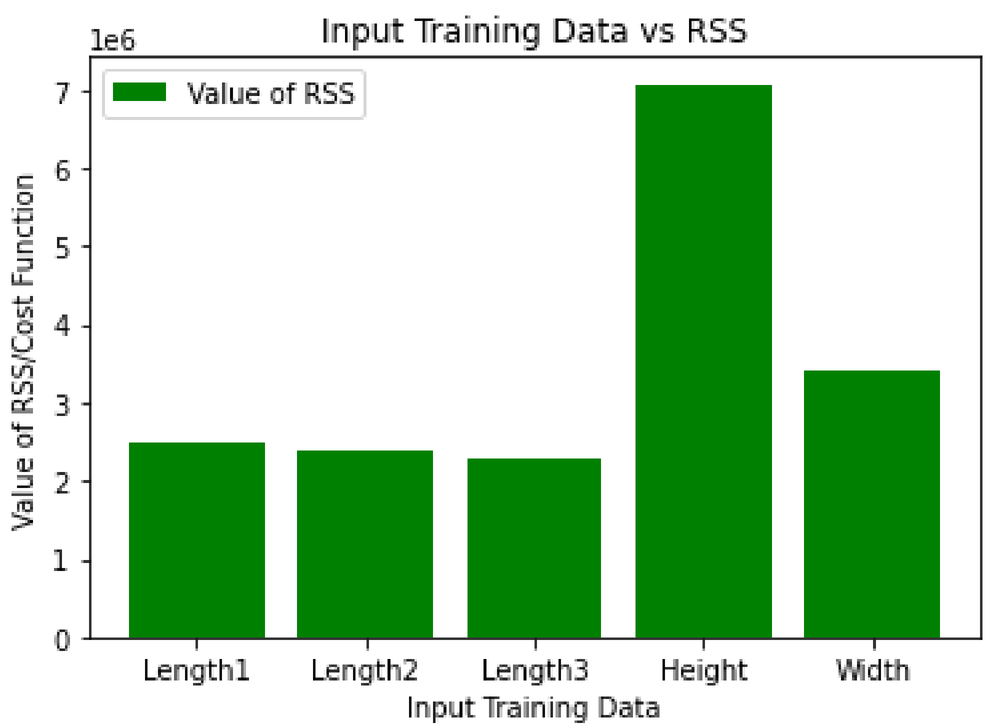


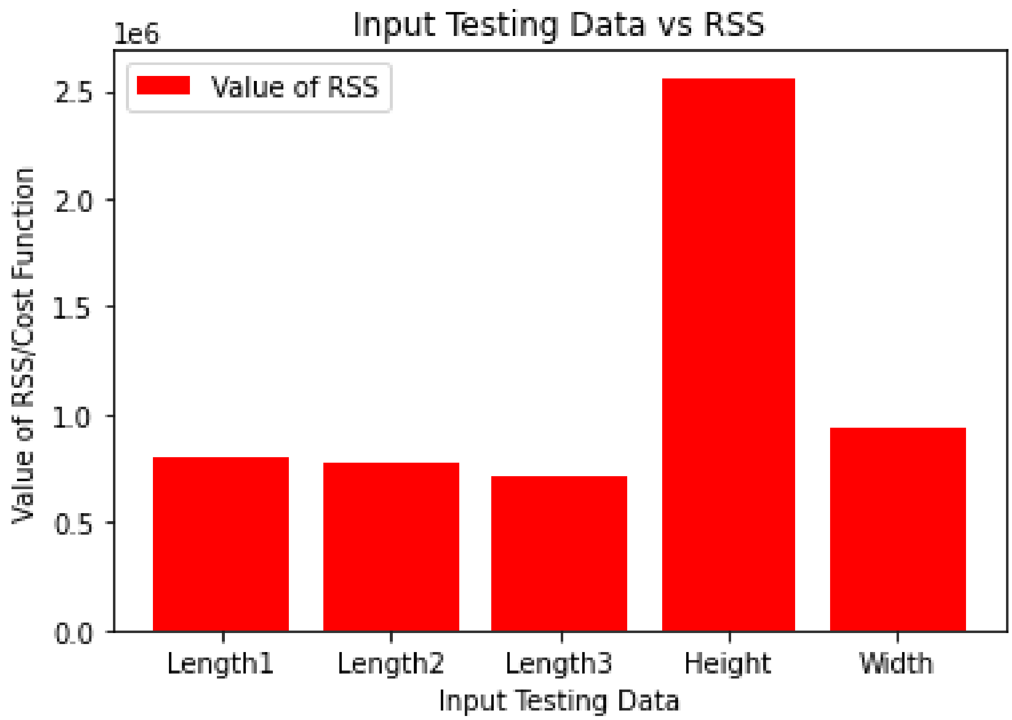
Length-2 vs Weight
Line passing through Predicted data
Actual Data Points
1500
1250
1000
750
500
250
0
−250
10
20
30
40
50
60
y_train['Weight']
x_train['Length-2']


Length-3 vs Weight
Line passing through Predicted data
Actual Data Points
1500
1250
1000
750
500
250
0
−250
10
20
30
40
50
60
y_train['Weight']
x_train['Length-3']


Height vs Weight
Line passing through Predicted data
Actual Data Points
1500
1250
1000
750
500
250
0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
y_train['Weight']
x_train['Height']


Width vs Weight
Line passing through Predicted data
Actual Data Points
1500
1250
1000
750
500
250
0
−250
y_train['Weight']
1
2
3
4
5
6
7
8
x_train['Width']


Input Training Data vs RSS
1e6
Value of RSS
7
6
5
4
3
2
1
0
Value of RSS/Cost Function
Length1
Length2
Length3
Height
Width
Input Training Data



Input Testing Data vs RSS
1e6
Value of RSS
2.5
2.0
1.5
1.0
0.5
0.0
Value of RSS/Cost Function
Length1
Length2
Length3
Height
Width
Input Testing Data