MACHINE LEARNING(CSCE 633)

Assignment – 01 Report

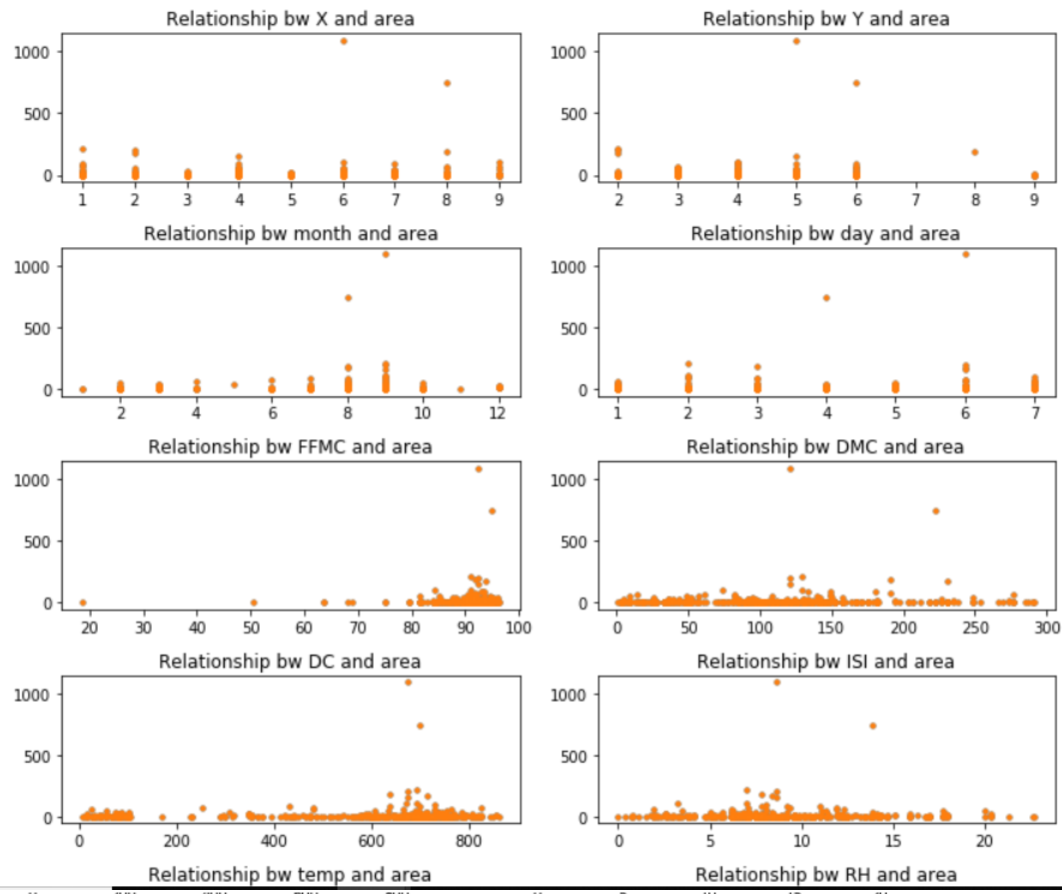
Amiya R Panda

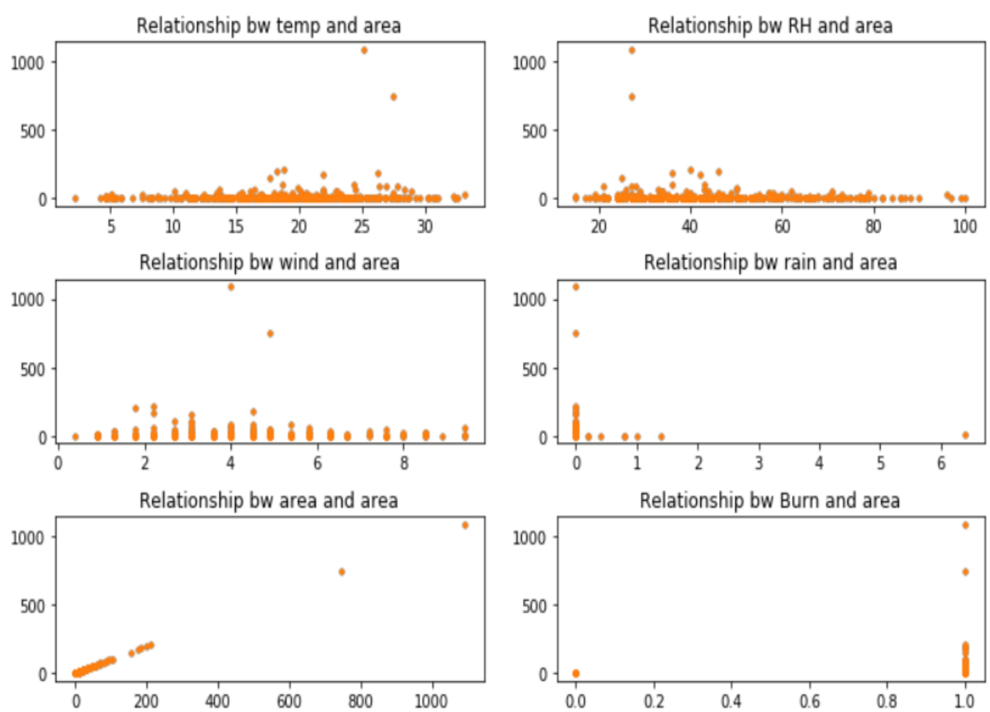
UIN – 727006179

Q3. **A. DATA EXPLORATION**

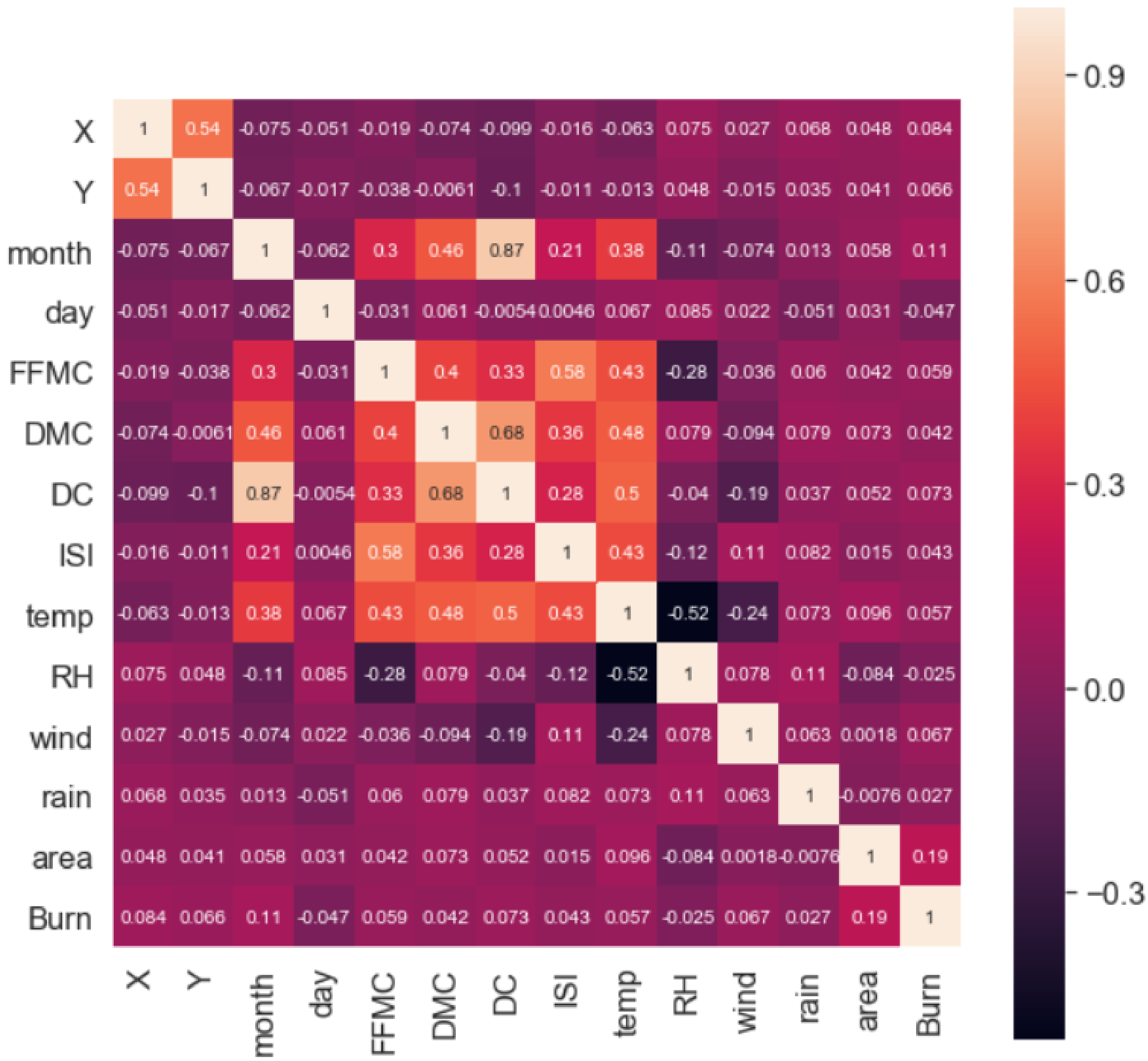
From the data exploration (scatter plot) we get the intuition of the distribution of different features in the dataset. We can classify the features into categorical and continuous types. Accordingly we can use different strategies for evaluating, normalizing, feature selection and so on. From the below given charts, we can conclude that the features "X", "Y", "month", and "day" are categorical features and the rest of them are continuous features. In this particular case the feature "rain" can also be treated as a categorical feature as it takes only between a few values.

Below given is the scatterplot for NOT NORMALISED DATA.

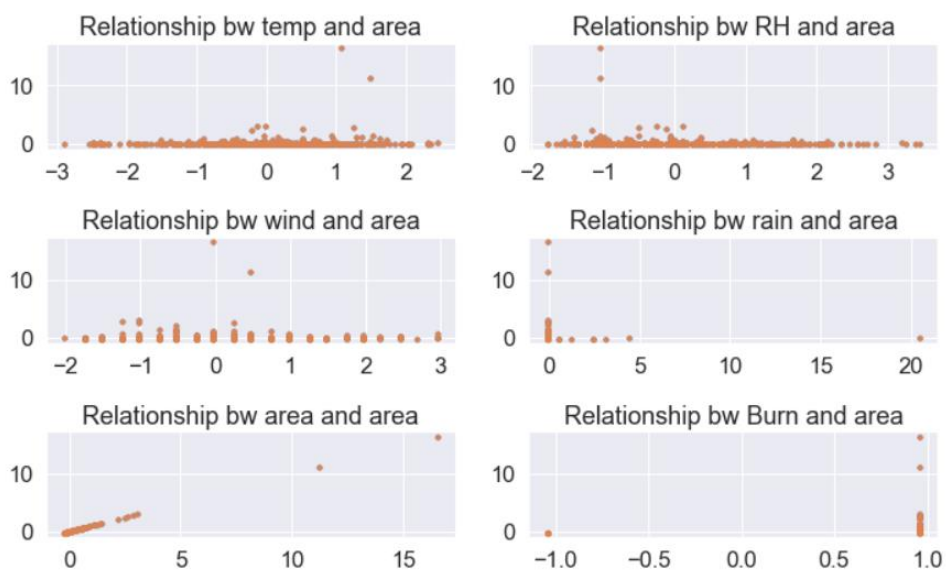

Relationship bw X and area
Relationship bw Y and area
Relationship bw month and area
Relationship bw day and area
Relationship bw FFMC and area
Relationship bw DMC and area
Relationship bw DC and area
Relationship bw ISI and area
Relationship bw temp and area
Relationship bw RH and area


Relationship bw temp and area
Relationship bw RH and area
Relationship bw wind and area
Relationship bw rain and area
Relationship bw area and area
Relationship bw Burn and area

Below given is the correlation between different features.



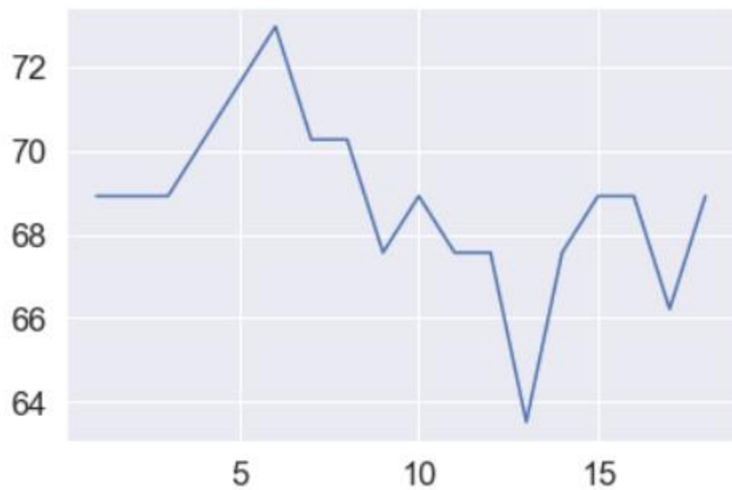
Below given the scatterplot for NORMALISED DATA(with Z score).



B. **KNN CLASSIFICATION**

Knn is implemented in a modular approach along with 5 fold cross validation module. Hyperparameter tuning is performed to find the value of "k" for which the result is optimum. The implementation considers EUCLIDEAN DISTANCE as the distance metric and is normalized using Z-SCORE technique.

(comments in the code mentions the functionality of the modules)



On X- Axis we have the value of k and on the Y – Axis the percentage of accuracy.

The values in the Y – Axis , represented in the vector form with index no. as the value of k is given below.

accuracy_train =

[68.91891891891892,

68.91891891891892,

68.91891891891892,

70.27027027027027,

71.62162162162163,

72.97297297297297,

70.27027027027027,

70.27027027027027,

67.56756756756756,

68.91891891891892,

67.56756756756756,

67.56756756756756,

63.51351351351351,

67.56756756756756,

68.91891891891892,

68.91891891891892,

66.21621621621621,

68.91891891891892]

We observed that at k = 6, the accuracy of the model is maximum that is 72.9%. So, we trained the model at k=6 and for the test set the accuracy is 64.51612903225806 %.

C. **LINEAR REGRESSION**

First the rows in the dataset is cut in places where the area burned is 0. These are irrelevant data for regression model. So, we are left with 243 samples.

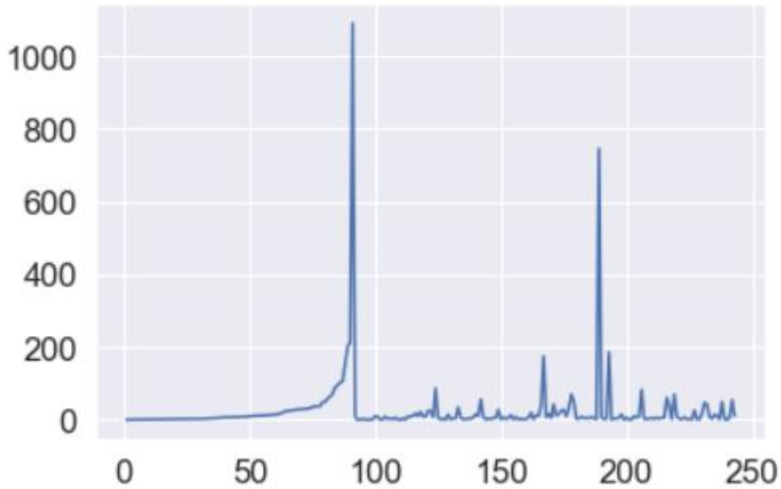
We can observe from the graphs that the data is distribution is uniform and has a very few outliers.

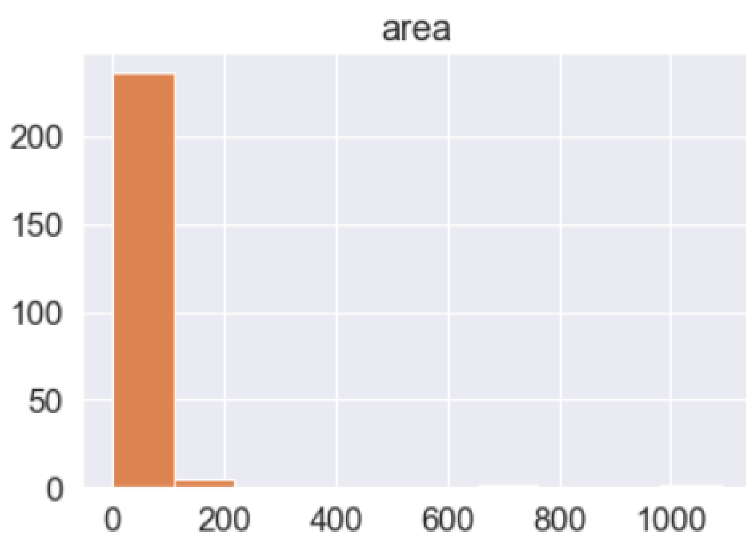
The histogram plot shows that the frequency of data is very high below the value of 100 and exponentially decreases going right onwards. The same is conspicuous from the log-scaled histogram.

The linear regression is implemented with OLS(analytical) technique without normalized data and the RSS value comes as 45303. (Detailed implementation mentioned in the comments.)

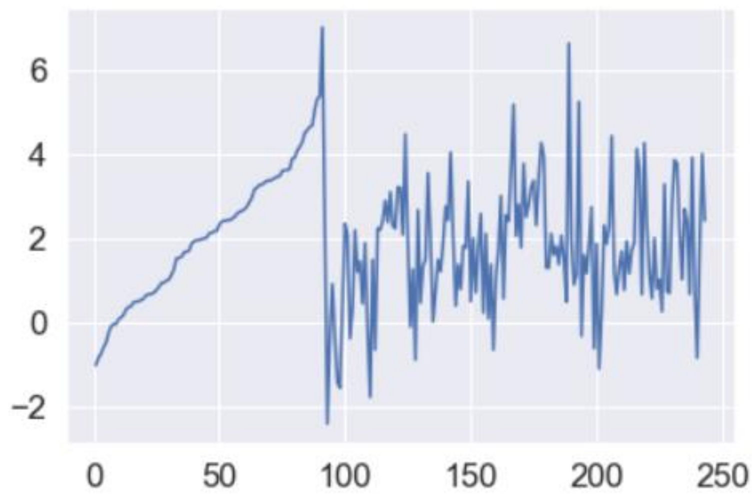
For normalized data, the model works very good and gives a significantly good result.

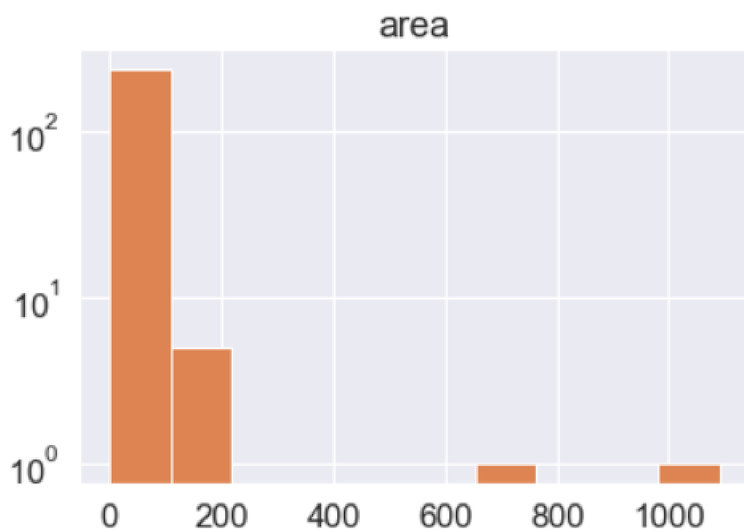
Below given is the data distribution graph for SAMPLE NO. VS AREA.(for line graph)





Below given graphs represents LOG(AREA) VS NO. OF SAMPLE.





The output weights of the model is given below

W = [[0.10777118],

[0.01402383],

[0.03082853],

[0.08156124],

[-0.01451973],

[0.1396619],

[-0.04999264],

[-0.0694056],

[0.06738026],

[-0.17265684],

[0.04232795],

[-0.00763402]]

The RSS error for train data is test TRAIN_RSS = 0.5707773115884475.

The RSS error for test data is TEST_RSS = 5.360540189872757.

The correlation between the model output and the actual output is

[[1. , -0.21727893],

[-0.21727893, 1.]]

The model output for the test data are

[-0.15336137, -0.20108913, 0.2423957 , -0.27320978, 0.08342409,

-0.05181747, 0.18448194, -0.12484288, 0.16579572, 0.27399939,

-0.08558736, -0.13581661, -0.29227719, 0.12952141, 0.08988106,

-0.42631901, -0.24293733, -0.00679973, -0.23183911, 0.37511037,

-0.08031317, 0.19571014, 0.01732945, 0.17538981, 0.27515074,

-0.13014322, 0.25087294, 0.2042629 , 0.3783526 , -0.25734019,

-0.17767834, -0.06246897, -0.15014932, -0.31692415, 0.07042788,

-0.15965736, -0.0634074 , 0.34648168, 0.24426665, 0.3588779 ,

0.42399978, -0.10801515, -0.27458979, -0.36763298, 0.05401277,

0.55357879, 0.14428057, -0.08695152, 0.26177797]

The actual output are

[-2.73034615e-01, -2.67179256e-01, -2.61436500e-01, -2.59522248e-01,

-2.56481965e-01, -2.55243331e-01, -2.49387972e-01, -2.18985146e-01,

-1.96914947e-01, -1.53337563e-01, -1.48045219e-01, -1.41626845e-01,

-1.30929554e-01, -8.41992849e-02, 4.48438207e-02, 1.38979978e-01,

2.78044756e-01, 8.86326480e-01, 2.11921737e+00, -2.70219538e-01,

-2.30020246e-01, -2.60310469e-01, -1.57053464e-01, 1.04123076e-03,

-1.16966775e-01, -2.60085263e-01, -2.28218597e-01, -2.66503637e-01,

-1.01765362e-01, -2.61323897e-01, -1.95563710e-01, -2.10652520e-01,

-1.86442862e-01, -2.44883850e-01, -2.14255818e-01, -9.85624306e-04,

-2.09864299e-01, -1.90046160e-01, -2.59522248e-01, -2.71795981e-01,

-2.64026370e-01, 6.53913767e-01, -2.42068774e-01, -2.56031553e-01,

-1.64372663e-01, -2.46910705e-01, -1.12575256e-01, -2.53441683e-01,

-2.05360176e-01]

NO1 (a.)

$$RSS(w_0, w_1) = \sum_{n=1}^{N} (y_n - w_0 - w_1 x_n)^2$$

$$= \sum_{n=1}^{N} (y_n^2 + w_0^2 + w_1^2 x_1^2 - 2y_n w_0 - 2 y_n w_1 x_n + 2 w_0 w_1 x_n)$$

$$\frac{\partial RSS}{\partial w_0} = \sum_{n=1}^{N} 2 \times (y_n - w_0 - w_1 x_n) \times -1 = 0$$

$$\Rightarrow \sum_{n=1}^{N} y_n = \sum_{n=1}^{N} w_0 + \sum_{n=1}^{N} w_1 x_n.$$

$$\Rightarrow N \times w_0 + w_1 \times \sum_{n=1}^{N} x_n = \sum_{n=1}^{N} y_n$$

$$\Rightarrow w_0 = \frac{1}{N} \times \sum_{n=1}^{N} y_n - \frac{w_1}{N} \times \sum_{n=1}^{N} x_n \quad \text{--- (1.)}$$

$$\frac{\partial RSS}{\partial w_1} = 0$$

$$\Rightarrow \sum_{n=1}^{N} 2 \times (-x_n)(y_n - w_0 - w_1 x_n) = 0$$

$$\Rightarrow \sum_{n=1}^{N} (x_n y_n - w_0 x_n - w_1 x_n^2) = 0$$

$$\Rightarrow -w_1 \sum_{n=1}^{N} x_n^2 + \sum_{n=1}^{N} x_n y_n - w_0 \sum_{n=1}^{N} x_n = 0$$

Subtituting eqn (1.)

$$\Rightarrow -w_1 \sum_{n=1}^{N} x_n^2 + \sum_{n=1}^{N} x_n y_n - \left[\frac{1}{N} \times \sum_{n=1}^{N} y_n - \frac{w}{N} \times \sum_{n=1}^{N} x_n\right] \times \sum_{n=1}^{N} x_n = 0$$

$$\Rightarrow -w_1\left[\sum_{n=1}^{N} x_n^2 - \frac{1}{N} \times \sum_{n=1}^{N}(x_n)\right] + \left[\sum_{n=1}^{N} x_n y_n - \frac{1}{N}\sum_{n=1}^{N} x_n \sum_{n=1}^{N} y_n\right] = 0$$

$$\Rightarrow \quad w_1 = \frac{\left[\sum_{n=1}^{N} x_n y_n - \frac{1}{N} \times \sum_{n=1}^{N} x_n \times \sum_{n=1}^{N} y_n\right]}{\sum_{n=1}^{N} x_n^2 - \frac{1}{N} \times \sum_{n=1}^{N} x_n.} \quad — (2.)$$

(b.) from eqn (1.)

$$w_0 = \frac{1}{N} \times \sum_{n=1}^{N} y_n - w_1 \times \frac{1}{N} \times \sum_{n=1}^{N} x_n$$

$$= (\overline{y_n} - w_1 \cdot \overline{x_n}) \quad — (3.)$$

from eqn (2.)

$$w_1 = \frac{\left[\sum_{n=1}^{N} x_n y_n - \frac{1}{N} \times \sum_{n=1}^{N} x_n \times \sum_{n=1}^{N} y_n\right]}{\sum_{n=1}^{N} x_n^2 - \frac{1}{N} \times \sum_{n=1}^{N} x_n.}$$

$$W_1 = \frac{\sum_{n=1}^{N} \left[(x_n - \bar{x})(y_n - \bar{y}) \right]}{\sum_{n=1}^{N} (x_n - \bar{x})^2}$$

$$= \frac{\sum_{n=1}^{N} \left[x_n y_n - x_n \bar{y} - y_n \bar{x} + \bar{x}\bar{y} \right]}{\sum_{n=1}^{N} \left(x_n^2 + \bar{x}^2 - 2 x_n \bar{x} \right)}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n - \bar{y} \sum_{n=1}^{N} x_n - \bar{x} \sum_{n=1}^{N} y_n + N \times \bar{x} \times \bar{y}}{\sum_{n=1}^{N} x_n^2 + (\bar{x})^2 \times N - 2 \bar{x} \times \sum_{n=1}^{N} x_n}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n - \bar{y} \cdot \bar{x} \cdot N - \bar{x} \cdot \bar{y} \cdot N + N \cdot \bar{x} \cdot \bar{y}}{\sum_{n=1}^{N} x_n^2 + (\bar{x})^2 \times N - 2 \times \bar{x} \times N \times \bar{x}}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n - \bar{y} \times \bar{x} \times N}{\sum_{n=1}^{N} x_n^2 - N \times (\bar{x})^2}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n - N \times \frac{1}{N} \times \sum_{n=1}^{N} x_n \times \frac{1}{N} \times \sum_{n=1}^{N} y_n}{\sum_{n=1}^{N} x_n^2 - N \times \left(\frac{1}{N} \times \sum_{n=1}^{N} x_n \right)^2}$$

(c) $w_0 = \underline{\text{mean of } \{y\}} - w_1 \times \underline{\text{mean of } \{x\}}$

$$w_1 = \frac{\sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y})}{\sum_{i=1}^{N} (x_n - \bar{x})^2}$$

$\bar{x}, \bar{y}$ → mean of $\{x_n\}$ and $\{y_n\}$

$$= \frac{\sum_{n=1}^{N} x_n y_n - N \times (\mu_x) \times (\mu_y)}{N \times (\sigma_x^2)}$$

where, $\mu_x \rightarrow$ mean of $\{x\}_1^N$

$\mu_y \rightarrow$ mean of $\{y\}_1^N$

$\sigma_x \rightarrow$ Variance of $\{x\}_1^N$

NO2 (a.) Taylor Series expansion.

$$f(x) = f(x_0) + \left(\nabla f\Big|_{x=x_0}\right)^T (x-x_0) + \frac{1}{2}(x-x_0)^T . H_f\Big|_{x=x_0} (x-x_0) \dots$$

for $f(x) = J(w)$, the taylor Series expansion will be Subtituting x with $w(K)$ at K^{th} iteration and $f(x)$ with $J(x)$, value of $w(K)$ is w at K^{th} iteration.

$$J(w) = J(w(k)) + \left(\nabla J\Big|_{w=w(k)}\right)^T (w - w(k)) + \frac{1}{2} \times (w-w(k))^T . H_J\Big|_{w=w(k)} (w-w(k)) + \dots$$

we ignore the lower order terms to say $J(w)$ approximately equals the Sum OF the zeroth, first and second term.

(b.) $$J(w(k+1)) \approx J(w(k)) + \left(\nabla J\Big|_{w=w(k)}\right)^T . (w(k+1) - w(k)) + \frac{1}{2}(w(k+1) - w(k))^T . H_J\Big|_{w=w(k)} (w(k+1) - w(k))$$

subtituting,

$$w(k+1) = w(k) - \alpha(k) . \nabla J\Big|_{w=w(k)}$$

$$J((w+1)) \approx J(w(k)) + \left(\nabla J\big|_{w=w(k)}\right)^T \times \left(\cancel{w(k)} - \alpha(k)\nabla J\big|_{w=w(k)} - \cancel{w(k)}\right) + \frac{1}{2}\left[\cancel{w(k)} - \alpha(k)\nabla J\big|_{w=w(k)} - \cancel{w(k)}\right]^T \times H_J\big|_{w=w(k)} \times \left(\cancel{w(k)} - \left(\alpha(k) \times \nabla J\big|_{w=w(k)}\right) - \cancel{w(k)}\right)^2$$

$$\Rightarrow \boxed{J(w(k+1)) \approx J(w) - \alpha(k)\left(\nabla J\big|_{w=w(k)}\right)^T \left(\nabla J\big|_{w=w(k)}\right) + \frac{1}{2}\left(-\alpha(k) \cdot \nabla J\big|_{w=w(k)}\right)\left(H_J\big|_{w=w(k)}\right)\left(-\alpha(k)\nabla J\big|_{w=w(k)}\right)}$$

$$= J(w) - \alpha(k)\left(\nabla J\big|_{w=w(k)}\right)^T\left(\nabla J\big|_{w=w(k)}\right) + \frac{1}{2} \times (\alpha(k))^2 \left(\nabla J\big|_{w=w(k)}\right)^T H_f\big|_{w=w(k)} \left(\nabla J\big|_{w=w(k)}\right) \qquad (1)$$

C

(c.) from eqⁿ (1)

$$J(w+1) = J(w(k)) - \left|\nabla J\big|_{w=w(k)}\right|^2 . \alpha(k) + \frac{1}{2}\left(\nabla J\big|_{w=w(k)}\right)^T . H_f\big|_{w=w(k)} \left(\nabla J\big|_{w=w(k)}\right) . \alpha^2(k)$$

Differentiating w.r.t. $\alpha(k)$

$$0 = 0 - \left|\nabla J\big|_{w=w(k)}\right|^2 + \left(\nabla J\big|_{w=w(k)}\right)^T . H_f\big|_{w=w(k)} \left(\nabla J\big|_{w=w(k)}\right) . \alpha(k)$$

$$\Rightarrow \alpha(k) = \frac{\left|\nabla J\big|_{w=w(k)}\right|^2}{\left(\nabla J\big|_{w=w(k)}\right)^T H_f\big|_{w=w(k)} \left(\nabla J\big|_{w=w(k)}\right)}$$