$$C = \frac{-1}{m} \sum_i \left(y_i \log(y_p) + (1-y_i)\log(1-y_p) \right)$$

$$\frac{\partial C}{\partial \beta_i} = \frac{-1}{m} \left(\frac{y_i}{y_p} - \frac{(1-y_i)}{(1-y_p)} \right) \frac{\partial y_p}{\partial \beta}$$

y_p = y predicted.

$$y_p = \sigma(\beta \cdot x)$$

$$\frac{\partial y_p}{\partial \beta} = \frac{\partial \sigma}{\partial (\beta \cdot x)} \frac{\partial \beta \cdot x}{\partial \beta}$$

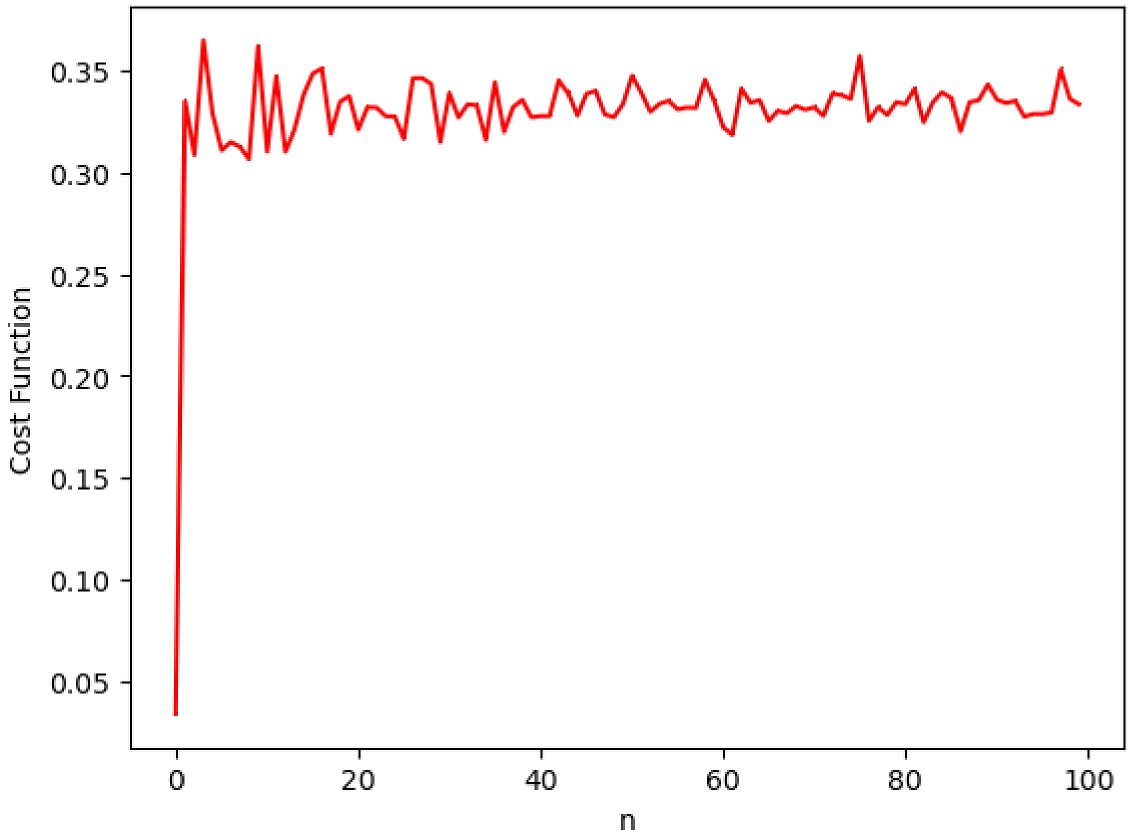
σ = sigmoid function

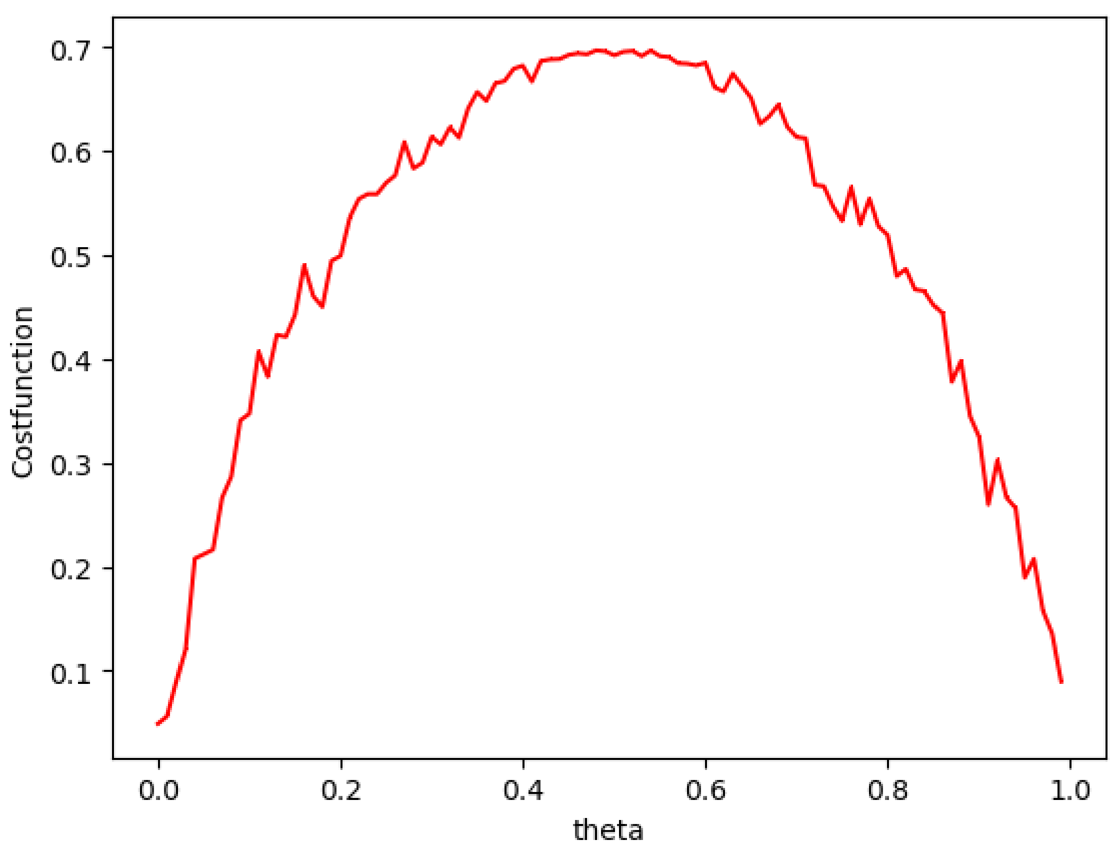
$$= \sigma(1-\sigma) \cdot x$$

$$= y_p(1-y_p)\, x.$$

$$\frac{\partial C}{\partial \beta_i} = \frac{-1}{m} \left(\frac{y_i(1-y_p) - y_p(1-y_i)}{y_p(1-y_p)} \right) \cdot y_p(1-y_p) \cdot x$$

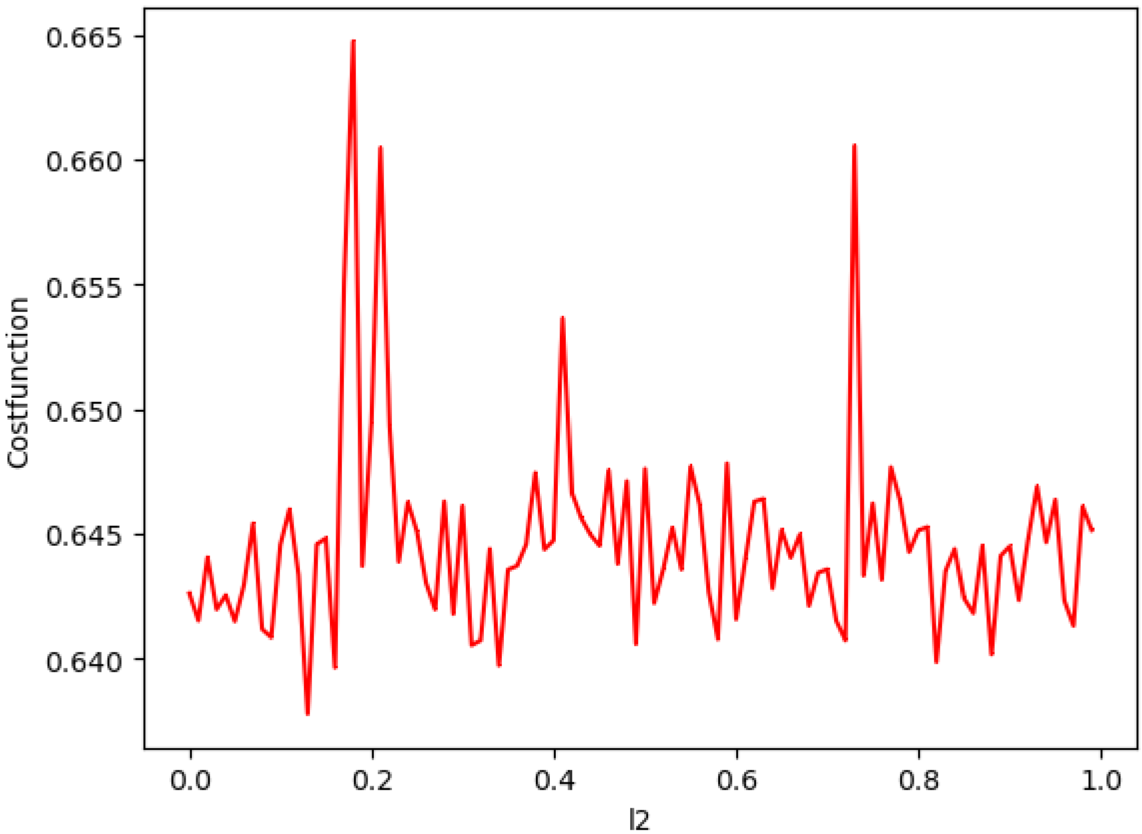
$$= \frac{1}{m}(y_p - y_i)\, x$$


0.35
0.30
0.25
0.20
0.15
0.10
0.05
Cost Function
0
20
40
60
80
100
n


0.7
0.6
0.5
0.4
0.3
0.2
0.1
Costfunction
0.0
0.2
0.4
0.6
0.8
1.0
theta

- As we increase n, the cost function spikes initially and as n increases progressively the cost function decreases converges to a value.
- As we increase theta, the cost function increases and reaches a maximum at theta = 0.5. This is because the data has the maximum noise at theta = 0.5. Then cost function decreases and reaches the same initial value at theta = 1.

Q4



- Cost function decreases with increasing values of lambdas(coefficients of l1 and l2 realization), for lambda between a certain range for different datasets. One can see them as dips in the cost function values. The sharp increases after these points implies underfitting and overfitting due of the model. It is important to tweak the values of lambda 1 and lambda 2 according to our model