1) Consider a logistic regression model with the feature vector $x = [1, 4, 2]$ and the true label $y = 1$. The model parameters are $\theta = [0.5, -0.3, 0.8]$, and the regularization parameter $\lambda = 0.5$. Calculate the cost function for logistic regression with L2 regularization.

Solution:

The cost function for logistic regression with L2 regularization is given by:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

where,

$h_\theta(x)$ = logistic (sigmoid) function

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

m = no. of training samples (here, $m = 1$)

λ = regularization parameter

Given,

$x = [1, 4, 2]$, $\theta = [0.5, -0.3, 0.8]$

$$\theta^T x = (0.5 \times 1) + (-0.3 \times 4) + (0.8 \times 2)$$
$$= 0.5 - 1.2 + 1.6$$
$$= 0.9$$

$$h_\theta(x) = \frac{1}{1+e^{-0.9}}$$

let, $e^{-0.9} \simeq 0.4066$

$$h_\theta(x) \simeq \frac{1}{1+0.4066}$$

$$= \frac{1}{1.4066} \simeq 0.711$$

Since $y = 1$,

$$J(\theta) = -\log h_\theta(x) + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

$$J(\theta) = -\log(0.711) + \frac{0.5}{2 \times 1} \left[(-0.3)^2 + (0.8)^2\right]$$

let, $\log(0.711) \approx -0.340$

$$J(\theta) = 0.340 + \frac{0.5}{2}(0.09 + 0.64)$$

$$= 0.340 + 0.25 \times 0.73$$

$$= 0.340 + 0.1825$$

$$= 0.5225$$

So, $J(\theta) \approx 0.523$

2) Given a polynomial regression model $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$, with the true label $y = 25$, and the feature vector $x = 3$, calculate the cost function with L2 regularization. The parameters are $\theta = [2, 3, -1]$, and the regularization parameter $\lambda = 0.4$

Solution:

The cost function for polynomial regression with L2 regularization is given by:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

where,

$h_\theta(x)$ = polynomial regression hypothesis

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$m = 1$, we have single data point

λ = regularization parameter

Given,

$\theta = [2, 3, -1]$, $x = 3$

$$h_\theta(3) = 2 + (3 \times 3) + (-1 \times 3^2)$$
$$= 2 + 9 - 9 = 2$$

Squared Error Term –

$$(h_\theta(x) - y)^2 = (2 - 25)^2 = (-23)^2 = 529$$

$$\frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2 = \frac{0.4}{2 \times 1} \left[(3)^2 + (-1)^2\right]$$
$$= \frac{0.4}{2} \times (9 + 1)$$
$$= 0.2 \times 10$$
$$= 2$$

$$J(\theta) = \frac{1}{2} \times 529 + 2$$

$$= 264.5 + 2$$

$$= 266.5$$

So, $J(\theta) = 266.5$

3) Consider a linear regression model $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ with initial parameters $\theta_0 = 1$, $\theta_1 = 0.3$, $\theta_2 = -0.5$, and $\theta_3 = 0.8$ and regularization parameter $\lambda = 0.2$. Given the training example $(x_1, x_2, x_3, y) = (2, -1, 4, 7)$, update the parameters θ_0, θ_1, θ_2, and θ_3 using gradient descent with a learning rate $\alpha = 0.1$ for 3 iterations.

Solution:

The cost function for linear regression with L2 regularization is:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

where,

$h_\theta(x)$ = hypothesis function

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$m = 1$

λ = regularization parameter

For each parameter θ_j, the gradient descent update rule

$$\theta_j := \theta_j - \alpha\left(\frac{dJ(\theta)}{d\theta_j}\right)$$

Partial derivatives –

$$\frac{dJ}{d\theta_0} = (h_\theta(x) - y)$$

$$\frac{dJ}{d\theta_j} = (h_\theta(x) - y)x_j + \frac{\lambda}{m}\theta_j, \text{ for } j \geq 1$$

Given,

$\theta_0 = 1$, $\theta_1 = 0.3$, $\theta_2 = -0.5$, $\theta_3 = 0.8$

$x_1 = 2$, $x_2 = -1$, $x_3 = 4$, $y = 7$

$\lambda = 0.2$, $\alpha = 0.1$

Iteration 1

$$h_\theta(x) = 1 + (0.3 \times 2) + (-0.5 \times -1) + (0.8 \times 4)$$

$$= 1 + 0.6 + 0.5 + 3.2$$

$$= 5.3$$

Error:

$$h_\theta(x) - y = 5.3 - 7 = -1.7$$

$$\frac{dJ}{d\theta_0} = -1.7$$

$$\frac{dJ}{d\theta_1} = (-1.7 \times 2) + \frac{0.2}{1} \times 0.3$$

$$= -3.4 + 0.06 = -3.34$$

$$\frac{dJ}{d\theta_2} = (-1.7 \times -1) + \frac{0.2}{1} \times (-0.5)$$

$$= 1.7 - 0.1 = 1.6$$

$$\frac{dJ}{d\theta_3} = (-1.7 \times 4) + \frac{0.2}{1} \times 0.8$$

$$= -6.8 + 0.16 = -6.64$$

$$\theta_0 = 1 - (0.1 \times -1.7) = 1.17$$

$$\theta_1 = 0.3 - (0.1 \times -3.34)$$

$$= 0.3 + 0.334 = 0.634$$

$$\theta_2 = -0.5 - (0.1 \times 1.6)$$

$$= -0.5 - 0.16 = -0.66$$

$$\theta_3 = 0.8 - (0.1 \times -6.64)$$

$$= 0.8 \times 0.664 = 1.464$$

Iteration 2

$h_\theta(x) = 1.17 + (0.634 \times 2) + (-0.66 \times -1) + (1.464 \times 4)$

$= 1.17 + 1.268 + 0.66 + 5.856$

$= 8.954$

Error:

$h_\theta(x) - y = 8.954 - 7 = 1.954$

$\frac{dJ}{d\theta_0} = 1.954$

$\frac{dJ}{d\theta_1} = (1.954 \times 2) + (0.2 \times 0.634)$

$= 3.908 + 0.1268$

$= 4.0348$

$\frac{dJ}{d\theta_2} = (1.954 \times -1) + (0.2 \times -0.66)$

$= -1.954 - 0.132$

$= -2.086$

$\frac{dJ}{d\theta_3} = (1.954 \times 4) + (0.2 \times 1.464)$

$= 7.816 + 0.2928$

$= 8.1088$

$\theta_0 = 1.17 - (0.1 \times 1.954) = 0.9746$

$\theta_1 = 0.634 - (0.1 \times 4.0348) = 0.2305$

$\theta_2 = -0.66 - (0.1 \times -2.086) = -0.4514$

$\theta_3 = 1.464 - (0.1 \times 8.1088) = 0.6531$

Iteration 3

$$h_\theta(x) = 0.9746 + (0.2305 \times 2) + (-0.4514 \times -1) + (0.6531 \times 4)$$
$$= 0.9746 + 0.461 + 0.4514 + 2.6124$$
$$= 4.4994$$

Error:

$4.4994 - 7 \approx -2.5006$

$\theta_0 \simeq 1.2247$

$\theta_1 \simeq 0.4805$

$\theta_2 \simeq -0.6513$

$\theta_3 \simeq 1.503$

So, the final values of θ_0, θ_1, θ_2 and θ_3 are 1.2247, 0.4805, -0.6513 and 1.503 respectively.