1. Linear regression aims to minimize the sum of squared differences between the actual values and the predicted values.

2. Linear regression can represent quadratic equations by including polynomial features. For instance, the equation $y = ax^2 + bx + c$ can be modelled using x and x^2 as input features.

3. Identifying and removing outliers is essential because they can distort statistical analyses and model predictions, leading to inaccurate results and reduced model performance.

4. Feature scaling involves normalizing the range of input variables. It is necessary when features have different scales, particularly for algorithms sensitive to data magnitude, such as k-nearest neighbours (KNN) and gradient descent methods.

5. Linear regression predicts continuous values, while logistic regression predicts binary outcomes. Linear regression uses least squares for optimization, whereas logistic regression relies on maximum likelihood estimation.

6. The Mean Square Error (MSE) cost function is not suitable for logistic regression because it can create non-convex optimization issues, complicating the search for the global minimum. Logistic regression uses the log-loss (cross-entropy) cost function, which is convex and better suited for binary classification.

7. If the cost function decreases initially but then increases or stagnates at a high value, it may indicate an excessively high learning rate causing overshooting, or the optimization process getting stuck in local minima or saddle points.

8. Train a separate binary logistic regression model for each class, treating that class as positive and others as negative. The class with the highest predicted probability is selected. Extend logistic regression to handle multiple classes using the SoftMax function to calculate probabilities for each class and choosing the class with the highest probability.

The mean squared error (MSE) is calculated as:

$$MSE\ (E(w)) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_w(x^{(i)}) - y^{(i)}\right)^2$$

$$h(x) = w_0 + w_1 x + w_2 x_2$$

initial weights:

$$w_0 = 0,\ w_1 = 1,\ w_2 = 1$$

For each data point:

$$h(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2$$
$$= 0 + 1\cdot x_1 + 1\cdot x_2$$

$$h(60, 22) = 60 + 22 = 82$$
$$h(67, 24) = 67 + 24 = 91$$
$$h(71, 15) = 71 + 15 = 86$$
$$h(75, 20) = 75 + 20 = 95$$
$$h(78, 16) = 78 + 16 = 94$$

$$MSE = \cancel{\frac{1}{10}} \frac{1}{2\times5}\left[(82-140)^2 + (91-159)^2 + (86-192)^2 + (95-200)^2 + (94-212)^2\right]$$

$$= 4327.3$$

$$w_j := w_j - \alpha \cdot \frac{\delta}{\delta w_j}(E(w))$$

Gradient for $w_0 = \frac{1}{5}\sum_{i=1} (h(x_i) - y_i)$

$$= \frac{1}{5}\left[(82-140) + (91-159) + (86-192) + (95-200) + (94-212)\right]$$

$$= -91$$

CamScanner

Gradient for $w_1 = \frac{1}{5}\sum_{i=1}^{5}(h(x_i) - y_i)x_{1i}$

$$= \frac{1}{5}\Big[(82-140)\times 60$$
$$+ (91-159)\times 67$$
$$+ (86-192)\times 71$$
$$+ (95-200)\times 75$$
$$+ (94-200)\times 78\Big]$$

$$= \frac{1}{5}\Big[(-58)\times 60 - (68\times 67)$$
$$-106\times 71 - 105\times 75$$
$$-118\times 78\Big]$$

$$= \frac{1}{5}\Big[-3480 - 4556 - 7526$$
$$-7875 - 9204\Big]$$

$$= \frac{-23641}{5} = -4728.2$$

$$\text{Gradient } w_2 = \frac{1}{5}\sum_{i=1}^{5}\left(h(x_i) - y_i\right)x_{2i}$$

$$= \frac{1}{5}\Big[(82-140)\times 22$$

$$+(91-159)\times 24$$

$$+(86-192)\times 15$$

$$+(95-200)\times 20$$

$$+(94-200)\times 16\Big]$$

$$= -1699.2$$

update weights for iteration 1:

$$w_0^{(1)} = w_0 - \alpha \times \text{Gradient for } w_0$$

$$= 0 - 0.0002 \times (-91)$$

$$= 0.0182$$

$w_1^{(1)} = w_1 - \alpha \times \text{Gradient for } w_1$

$= 1 - 0.0002 \times (-4228.2)$

$= 1 + 0.9456$

$= 1.9456$

$w_2^{(1)} = w_2 - \alpha \times (\text{Gradient for } w_2)$

$= 1 - 0.0002 \times (-1699.2)$

$= 1.3398$

Iteration 2

Using updated weights –

$w_0 = 0.0182$, $w_1 = 1.9456$ and $w_2 = 1.3398$

$h(60, 22) = 0.0182 + 1.9456 \times 60 + 1.3398 \times 22$

$= 146.2302$

$h(67, 24) = 0.0182 + 1.9456 \times 67$

$+ 1.3398 \times 24$

$= 0.0182 + 130.3552$

$+ 32.1552$

$= 162.9026$

$h(71, 15) = 0.0182 + 1.9456 \times 71$

$+ 1.3398 \times 15$

$= 158.4548$

$h(75, 20) = 0.0182 + 1.9456 \times 75$

$+ 1.3398 \times 20$

$= 171.7342$

$h(78, 16) = 0.0182 + 1.9456 \times 78$

$+ 1.3398 \times 16$

$= 172.6038$

CamScanner

Gradients for iteration 2 :

$$\text{Gradient for } w_0 = \frac{1}{5}\sum_{i=1}^{5}(h(x_i)-y_i)$$

$$= \frac{1}{5}\Big[(146.2302-140)$$

$$+(162.9026-159)$$

$$+(158.4548-192)$$

$$+(171.7342-200)$$

$$+(172.6038-212)\Big]$$

$$= -18.01488$$

$$\text{Gradient for } w_1 = -247.35328$$

$$\text{Gradient for } w_2 = -293.66516$$

$$w_0^{(2)} = w_0^{(1)} - \alpha \times \text{Gradient for } w_0$$

$$= 0.0182 - 0.0002 \times (-18.01488)$$

$$= 0.0218$$

CamScanner

$$w_1^{(2)} = w_1^{(1)} - \alpha \times \text{Gradient for } w_1$$

$$= 1.9456 - 0.0002 \times (-247.35328)$$

$$= 1.9951$$

$$w_2^{(2)} = w_2^{(1)} - \alpha \times \text{Gradient for } w_2$$

$$= 1.3398 - 0.0002 \times (-293.6651)$$

$$= 1.3985$$

$$MSE = \frac{1}{10}\Big[(146.2302 - 140)^2 + (162.9026 - 159)^2$$

$$+ (158.4548 - 192)^2$$

$$+ (171.7342 - 200)^2$$

$$+ (172.6038 - 212)^2\Big]$$

$$= 254.16425$$

	w_0	w_1	w_2
After iterate 1	0.0182	1.9456	1.3398
After iter 2	0.0218	1.9951	1.3985

Initial Mean Square error ⇒ 4327.3

Final Mean square error ⇒ 259.16425