Question 1

a. What does linear regression try to optimize?

Answer: Linear regression tries to optimize, using techniques like Gradient descent, the values of models parameters w0, w1 so that the cost function (MSE) is at the minimum, which helps in models accuracy and performance.

b. Is it possible to use linear regression to represent quadratic equations? Explain with an example.

Answer: Yes, it is possible to use linear regression to represent quadratic equations. The quadratic equations are of the form $y= ax^2+bx+c$. The linear regression can also be used if the line fitting the data is not straight line but with a curve. It also can be used equations where predicted value $y' = w0+ w1x +w2x^2$.

c. Why is it crucial to detect and remove outliers?

Answer: The outliers are extreme or unusual values out the normal range of data. The outliers affect the accuracy of linear regression. The outliers affect the slope and intercept of the line and also the variance of the error. The outliers also affect the relationship between the variables and create incorrect relationships

d. What is feature scaling? When is it required?

Answer: When some of the features are in different range of values sometimes differences are big, it can make some features dominant in the calculations, we have to make sure the features are on same scale and given equal weightage. We can use normalization strategies to normalize the feature values.

e. State two differences between linear regression and logistic regression.

Answer: Linear regression is used for solving regression problems and logistic regression is used for solving classification problems. The cost function used for Linear regression is Mean square Error(MSE) and cross entropy is used as cost function for Logistic regression

f. Why is the Mean Square Error cost function unsuitable for logistic regression?

Answer: For gradient descent analysis, we need to use cost function such that there is one minima and which helps in accurately predicting the optimum parameters with minimum cost function. The cost function should be convex. In case of logistic regression, mean square error is not convex and can result in multiple local minima. To avoid that, we need to use cost function like cross entropy which is convex in case of logistic regression

g. What can be inferred if the cost function initially decreases but then increases or gets stuck at a high value?

Answer: It means that the gradient descent is not working well. In case of convex cost function, it could be due to learning rate is too high and learning rate needs to be reduced

h. Describe two ways to perform multi-class classification using logistic regression.

Answer: There are two ways to perform multi-class classification using logistic regression.

1.One-vs-all or one-vs-rest classifier: We have to create separate classifiers for each class versus others

2.One-vs-one classifiers separate classifiers for each pair of classes

Question 2

Consider a linear regression model with two variables: h(x) = w0 + w1.x1 + w2.x2; which has been initialized with the following weights: w0 = 0; w1 = 1; w2 = 1. Consider the learning rate alpha = 0.0002. You are given the following data

x1	60	67	71	75	78
x2	22	24	15	20	16
y	140	159	192	200	212

Write the values of the weights after performing the gradient descent algorithm for **2 iterations**. Calculate the initial mean squared error before any iterations, and the final error after updating the weights for 2 iterations. Provide the values in tables like the following:

Answer:

	w0	w1	w2
After iteration 1	0.001	1.13	1.03
After Iteration 2	0.003	1.24	1.06

Initial Mean squared error	8834.6
Final Mean squared error	7111.5