Question 1:

a. What does linear regression try to optimize?
To determining best parameter (slope and Intercept) for model such that loss function of the model decreases as result of which model can predict more accurately.

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

Linear regression can't directly represent quadratic equation, as it models a linear relationship between variables. But quadratic data can sometimes be approximated using linear regression, especially if only part of the quadratic curve is considered.
Example:
Y=ax^2+bx+c, you would need to transform the features to include higher-order terms to use linear regression.

c. Why is it crucial to detect and remove outliers?
   i. outliers affect mean, standard deviation and regression line.
   ii. It affects Skew results and lead to incorrect conclusions
   iii. Predictive modeling performance is not accurate

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

Feature scaling is the process of standardizing the range of independent features of data. It requires when features have different scales, preventing the model from converging efficiently.

e. State two differences between linear regression and logistic regression

Linear Regression	Logistic Regression
It is suited for predicting continuous outcome	It is used for binary classification tasks with categorical outcome
Linear regression is based on the LSM (least square estimation)	Logistic regression is based on MLE (maximum likelihood estimation)

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

Mean Square Error cost function unsuitable for logistic regression because the logistic function used in logistic regression leads to non-convex cost function when squared, when causing optimization problem like local minima.

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

It indicates that the optimization process might be trapped in local minimum, or the learning rate is too high.

h. Describe two ways to perform multi-class classification using logistic regression
i. One-vs-all / one-vs-rest classifier: Separate classifier for each class vs others
ii. One-vs-one classifier: separate classifier for each pair of classes

Question 2:

	w0	w1	w2
After Iteration 1	0.0182	2.30564	1.33944
After Iteration 2	0.01674	2.2222	1.3004

Initial Mean squared error	4417.3
Final Mean squared error	135.62