## Question 1

Consider a binary classification problem where the response variable $y$ takes values in $\{0, 1\}$, and we aim to model the probability that $y = 1$ given the input features $X$. Logistic regression models the probability as:

$$P(y = 1|X) = \frac{1}{1 + \exp(-X\beta)}$$

where $X$ is the design matrix (with dimensions $n \times p$) and $\beta$ is the vector of coefficients. The log-likelihood function for logistic regression is:

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \left[ y_i \log\left(\frac{1}{1 + \exp(-X_i\beta)}\right) + (1 - y_i) \log\left(1 - \frac{1}{1 + \exp(-X_i\beta)}\right) \right]$$

1. Derive the gradient vector of the log-likelihood function with respect to $\beta$.
2. Show how the gradient descent algorithm can be used to update the coefficient estimates $\beta$ iteratively.

## Question 2

Implement a logistic regression model for binary classification using Python. We will use the **Titanic Survival Dataset**, which is available on Kaggle and other sources.

1. **Download the dataset:** - The dataset can be found at: Kaggle Titanic Dataset. - Alternatively, search for the Titanic dataset in CSV format and download it.
2. **Apply Logistic Regression:**
   - Perform basic preprocessing, including handling missing values and encoding categorical variables.
   - Split the data into a training set and a test set.
   - Fit a logistic regression model using Python's `sklearn` library.
   - Use cross-validation to evaluate the model's performance.
3. **Evaluate the Model:**
   - Report the accuracy, precision, recall, and F1-score on both the training and test sets.
   - Plot the Receiver Operating Characteristic (ROC) curve and calculate the Area Under the Curve (AUC).