1)Explain the difference between linear regression and logistic regression models. Provide an example of a scenario where logistic regression would be more appropriate.

Ans- Linear Regression is used to handle regression problems whereas Logistic regression is used to handle the classification problems. Linear regression provides a continuous output but Logistic regression provides discreet output

Linear regression is used to predict the continuous dependent variable using a given set of independent variables. Logistic Regression is used to predict the categorical dependent variable using a given set of independent variables. Linear Regression is used for solving Regression problem.

Logistic regression is used to calculate the probability of a binary event occurring, and to deal with issues of classification. For example, predicting if an incoming email is spam or not spam, or predicting if a credit card transaction is fraudulent or not fraudulent.

2) What is the cost function used in logistic regression, and how is it optimized?

Ans- A Cost function is used to gauge the performance of the Machine Learning model. A Machine Learning model devoid of the Cost function is futile. Cost Function helps to analyze how well a Machine Learning model performs. A Cost function basically compares the predicted values with the actual values.

3) Explain the concept of regularization in logistic regression and how it helps prevent overfitting.

Ans- Regularization is a way of finding a good bias-variance tradeoff by tuning the complexity of the model. It is a very useful method to handle collinearity (high correlation among features), filter out noise from data, and eventually prevent overfitting.

Regularization is a technique that penalizes the coefficient. In an overfit model, the coefficients are generally inflated. Thus, Regularization adds penalties to the parameters and avoids them weigh heavily. The coefficients are added to the cost function of the linear equation.

4) What is the ROC curve, and how is it used to evaluate the performance of the logistic regression model?

Ans- The ROC curve is produced by calculating and plotting the true positive rate against the false positive rate for a single classifier at a variety of thresholds. For example, in logistic regression, the threshold would be the predicted probability of an observation belonging to the positive class.

ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1.

5) What are some common techniques for feature selection in logistic regression? How do these

techniques help improve the model's performance?

Ans-

Chi-square Test.

Fisher's Score.

Correlation Coefficient.

Dispersion Ratio.

Backward Feature Elimination.

Recursive Feature Elimination.

Random Forest Importance.

6) How can you handle imbalanced datasets in logistic regression? What are some strategies for dealing with class imbalance?

Ans- In case be unbalanced label distribution, the best practice for weights is to use the inverse of the label distribution. In our set, label distribution is 1:99 so we can specify weights as inverse of label distribution. For majority class, will use weight of 1 and for minority class, will use weight of 99.

Techniques to Handle Imbalanced Data.

Use the right evaluation metrics.

Resample the training set.

Use K-fold Cross-Validation in the Right Way.

Ensemble Different Resampled Datasets.

Resample with Different Ratios.

Cluster the abundant class.

Design Your Models.

7) Can you discuss some common issues and challenges that may arise when implementing logistic regression, and how they can be addressed? For example, what can be done if there is multicollinearity among the independent variables?

Ans- Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. An example of logistic regression could be applying machine learning to determine if a person is likely to be infected with COVID-19 or not.

Some of the real-world examples where logistic regression models can be used are: Predict whether or not a customer will default on a loan. Predict whether or not a patient will have a heart attack. Predict whether or not an email is a spam.