Task 4 - Logistic Regression Interview Q&A

1. How does logistic regression differ from linear regression?

Linear regression predicts continuous values (e.g., price).

Logistic regression predicts probabilities between 0 and 1 for classification.

Example: Linear predicts house price; logistic predicts if an email is spam.

2. What is the sigmoid function?

Sigmoid is a mathematical function used in logistic regression to convert any number into a probability between 0 and 1.

Formula: sigmoid(z) = 1 / (1 + e^(-z))

3. What is precision vs recall?

Precision = TP / (TP + FP): How many predicted positives were correct.

Recall = TP / (TP + FN): How many actual positives were correctly predicted.

Precision focuses on quality; recall focuses on completeness.

4. What is the ROC-AUC curve?

ROC curve plots True Positive Rate vs False Positive Rate.

AUC (Area Under Curve) measures how well the model distinguishes between classes.

AUC close to 1.0 means excellent performance.

5. What is the confusion matrix?

A 2x2 table that compares actual vs predicted values:

TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative).

Helps evaluate classification performance.

6. What happens if classes are imbalanced?

The model may favor the majority class, leading to misleading accuracy.

Use metrics like precision, recall, and AUC.

Try resampling or adjusting class weights to fix it.

7. How do you choose the threshold?

Default threshold is 0.5, but it can be tuned based on the problem.

Lower threshold increases recall, higher threshold increases precision.

Use ROC or precision-recall curve to decide.

8. Can logistic regression be used for multi-class problems?

Yes. Logistic regression can be extended for multi-class classification using One-vs-Rest or softmax (multinomial).

Example: Classifying handwritten digits from 0 to 9.