Logistic Regression :

ASSIGNMENT QUESTIONS:

1. What is Logistic Regression, and how does it differ from Linear Regression?

- **Logistic Regression:** A statistical model used for binary classification. It predicts the probability of a binary outcome (e.g., yes/no, true/false, 0/1).
- **Linear Regression:** A statistical model used for predicting a continuous outcome (e.g., house price, temperature).
- **Key Difference:** Linear Regression outputs a continuous value, while Logistic Regression outputs a probability (between 0 and 1) that is then used for classification.

2. What is the mathematical equation of Logistic Regression?

The equation is:

- $p(y=1|x) = 1 / (1 + e^{(-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta n x n))})$

Where:

- $p(y=1|x)$ is the probability of the outcome being 1 given the input features x.
- $\beta_0, \beta_1, \beta_2, ..., \beta n$ are the coefficients (parameters) of the model.
- $x_1, x_2, ..., xn$ are the input features.
- e is Euler's number (approximately 2.71828).

3. Why do we use the Sigmoid function in Logistic Regression?

- The sigmoid function (1 / (1 + e^(-z))) transforms any real number into a value between 0 and 1, representing a probability.

- It maps the linear combination of features to a probability, making it suitable for binary classification.

4. What is the cost function of Logistic Regression?

The cost function is the **log loss** (or cross-entropy loss):

- $J(\beta) = - (1/m) \Sigma [y_i \log(p(y_i=1|x_i)) + (1-y_i) \log(1-p(y_i=1|x_i))]$

Where:

- m is the number of training examples.
- y_i is the actual label (0 or 1).
- $p(y_i=1|x_i)$ is the predicted probability.

5. What is Regularization in Logistic Regression? Why is it needed?

- **Regularization:** A technique used to prevent overfitting by adding a penalty term to the cost function.
- **Why it's needed:** Overfitting occurs when the model learns the training data too well, leading to poor performance on unseen data. Regularization constrains the model's coefficients, making it less sensitive to noise in the training data.

6. Explain the difference between Lasso, Ridge, and Elastic Net regression.

- **Ridge Regression (L2 Regularization):** Adds the squared magnitude of coefficients to the cost function. It shrinks coefficients towards zero but rarely sets them exactly to zero.
- **Lasso Regression (L1 Regularization):** Adds the absolute magnitude of coefficients to the cost function. It can shrink some coefficients to exactly zero, performing feature selection.
- **Elastic Net Regression:** A combination of L1 and L2 regularization. It adds both the squared and absolute magnitudes of coefficients. It balances the strengths of both Ridge and Lasso.

7. When should we use Elastic Net instead of Lasso or Ridge?

- Elastic Net is useful when you have a large number of features, and you suspect that some of them are correlated.
- It provides a balance between feature selection (Lasso) and coefficient shrinkage (Ridge).
- If you have many correlated features, lasso can randomly pick one, where as Elastic Net will pick groups of correlated features.

8. What is the impact of the regularization parameter (λ) in Logistic Regression?

- λ (lambda) controls the strength of the regularization.
- **Higher λ:** Stronger regularization, smaller coefficients, simpler model (less prone to overfitting).
- **Lower λ:** Weaker regularization, larger coefficients, more complex model (potentially prone to overfitting).
- If lambda is zero, then there is no regularization.

9. What are the key assumptions of Logistic Regression?

- **Binary Outcome:** The dependent variable must be binary.
- **Independence of Errors:** The observations should be independent of each other.
- **Linearity of Logits:** The log-odds of the outcome are linearly related to the predictor variables.
- **No Multicollinearity:** The predictor variables should not be highly correlated with each other.
- **Sufficiently Large Sample Size:** The sample size should be large enough to ensure stable parameter estimates.

10. What are some alternatives to Logistic Regression for classification tasks?

- Support Vector Machines (SVMs)
- Decision Trees
- Random Forests
- Gradient Boosting Machines (e.g., XGBoost, LightGBM)
- Neural Networks

11. What are Classification Evaluation Metrics?

- **Accuracy:** Overall correctness of predictions.
- **Precision:** Proportion of correctly predicted positives out of all predicted positives.
- **Recall (Sensitivity):** Proportion of correctly predicted positives out of all actual positives.
- **F1-score:** Harmonic mean of precision and recall.
- **AUC-ROC:** Area under the Receiver Operating Characteristic curve, which measures the model's ability to distinguish between classes.
- **Confusion Matrix:** A table that summarizes the model's predictions.

12. How does class imbalance affect Logistic Regression?

- Class imbalance occurs when one class has significantly more samples than the other.
- It can lead to biased models that favor the majority class, resulting in poor performance on the minority class.
- It causes the model to have higher accuracy on the majority class, and lower accuracy on the minority class.
- Techniques to handle this include oversampling the minority class, undersampling the majority class, and using weighted loss functions.

13. What is Hyperparameter Tuning in Logistic Regression?

- Hyperparameter tuning involves finding the optimal values for parameters that are not learned by the model (e.g., λ, solver, penalty).
- Techniques include grid search, random search, and Bayesian optimization.

14. What are different solvers in Logistic Regression? Which one should be used?

- **'liblinear':** Suitable for small datasets. Uses coordinate descent. Good for L1 regularization.
- **'lbfgs':** Suitable for small to medium datasets. Uses quasi-Newton methods. Good for L2 regularization.
- **'newton-cg':** Similar to 'lbfgs'. Good for L2 regularization.
- **'sag':** Suitable for large datasets. Uses stochastic average gradient descent.
- **'saga':** Suitable for large datasets. Supports L1, L2, and Elastic Net regularization.
- **Choice:** For small datasets, 'liblinear' or 'lbfgs'. For large datasets, 'sag' or 'saga'. 'saga' is versatile.

15. How is Logistic Regression extended for multiclass classification?

- **One-vs-Rest (OvR) or One-vs-All:** Train a binary Logistic Regression classifier for each class against all other classes.
- **Softmax Regression (Multinomial Logistic Regression):** Directly extends Logistic Regression to handle multiple classes by predicting the probabilities of each class.

16. What are the advantages and disadvantages of Logistic Regression?

- **Advantages:**
    - Simple and easy to implement.
    - Efficient to train.
    - Provides interpretable coefficients.
    - Works well for linearly separable data.

- **Disadvantages:**
    - Assumes linearity between features and log-odds.
    - Performs poorly on non-linear data.
    - Sensitive to outliers.
    - Can suffer from multicollinearity.

17. What are some use cases of Logistic Regression?

- Medical diagnosis (e.g., predicting disease risk).
- Credit risk assessment.
- Spam detection.
- Customer churn prediction.
- Marketing campaign success prediction.

18. What is the difference between Softmax Regression and Logistic Regression?

- **Logistic Regression:** Binary classification (two classes).
- **Softmax Regression:** Multiclass classification (multiple classes).
- Softmax outputs a probability distribution across all classes, while logistic regression outputs a probability of one class.

19. How do we choose between One-vs-Rest (OvR) and Softmax for multiclass classification?

- **OvR:** Simpler to implement, suitable when classes are not mutually exclusive.
- **Softmax:** More theoretically sound, suitable when classes are mutually exclusive.
- If the classes are truly exclusive, softmax is the better choice.

20. How do we interpret coefficients in Logistic Regression?

- Coefficients represent the change in the log-odds of the outcome for a one-unit change in the predictor variable, holding other variables constant.

- To get the odds ratio, exponentiate the coefficient (e^coefficient).
- A positive coefficient means an increase in the predictor variable increases the log-odds of the outcome, and vice-versa.