**Logistic Regression using Newton-Raphson Method**

Name: S. v. Syam Naga Pavan

**Libraries Needed:**

* ucimlrepo
* numpy
* pandas
* sklearn

**Dataset Retrieval:**

For the Breast Cancer Diagnostic dataset we use the fetch\_ucirepo function from ucimlrepo. This dataset contains features related to breast cancer cases, and the goal is to classify tumors as either malignant (M) or benign (B).

**Data Preprocessing:**

The dataset is preprocessed by mapping the target variable in the training set to the labels ‘M' (Malignant) to 1 and 'B' (Benign) to 0. This transformed the target variable into binary format, suitable for logistic regression.

**Data Splitting:**

The data is split into training, validation, and test sets. The training set comprises 80% of the original data, and validation and test sets account for 80% of the remaining data, ensuring no data leakage between these sets.

**Implementation of Logistic Regression**

We used the Newton-Raphson method, an iterative optimization approach, to implement logistic regression. The following are the essential elements of the logistic regression implementation:

**Sigmoid Function:** A probability value between 0 and 1 was obtained by applying the sigmoid function to the linear combination of features.

**Probability Calculation:** Using the sigmoid function, we determined the odds that the samples would belong to the positive class (Malignant).

**Weight Matrix:** We computed a weight matrix based on the probabilities to include it in the Hessian matrix computation.

**Hessian Matrix:** The features and the weight matrix were used to calculate the Hessian matrix. It is an essential part of the coefficient estimation process.

**Gradient Calculation:** The gradient was determined by multiplying the characteristics, targets, and probabilities by a dot product. In the parameter space, it denotes the direction of the sharpest ascent.

**Newton-Raphson Step:** The step when fresh coefficients are computed is the central component of logistic regression utilizing the Newton-Raphson technique. Additionally, regularization was applied, enabling the complexity of the model to be adjusted.

**Convergence Check:** To ensure the algorithm converges, a function was constructed to check for convergence based on the change in coefficients between iterations.

**Evaluation of correctness:** A function was developed to assess the correctness of the model. It computes the accuracy based on binary predictions made from probabilities.

**Results of Classification:**

Using the above hyperparameters, the Newton-Raphson method was used to train the logistic regression model. The subsequent outcomes were attained:

A maximum of 20 iterations

Following training, three distinct datasets were used to assess the model:

**Test Set Evaluation:** The model obtained a 95.65% accuracy on the test set after 20 iterations.

**Training Set Evaluation:** The model also attained a 96.70% accuracy on the training set.

**Validation Set Evaluation:** Throughout the training, the model's performance on the validation set was monitored, and after the prescribed number of iterations, it achieved an accuracy of 97–80%.

According to the classification results, the logistic regression model effectively distinguished between benign and malignant breast cancer cases. The accuracy of the training and test sets demonstrates the model's ability to generalize to new, unseen data. The validation set accuracy indicates that overfitting has been avoided and the model has been tweaked adequately. These findings make the Newton-Raphson method's logistic regression implementation a useful diagnostic tool for breast cancer. The model's performance can be enhanced by additional optimization and fine-tuning, which might be used in clinical settings.