**BREAST CANCER DIAGNOSIS USING LINEAR REGRESSION**

Implementation of Logistic Regression for classifying breast cancer as malignant or benign based on the Kaggle Breast Cancer Wisconsin Dataset.

**Overview:**

1. **Dataset**:
   * The dataset contains 569 samples with 32 features related to cell nuclei's characteristics.
   * Features are normalized for better performance.
   * The labels are encoded: "M" (Malignant) as 1 and "B" (Benign) as 0.
2. **Key Steps**:
   * **Data Preprocessing**:
     + Columns like id and Unnamed: 32 are dropped as they don't contribute to predictions.
     + Features are normalized using min-max scaling.
   * **Train-Test Split**:
     + Data is split into 85% training and 15% testing sets.
   * **Logistic Regression**:
     + Implements logistic regression from scratch with functions for:
       - **Initialization** of weights and bias.
       - **Sigmoid Function** for the hypothesis.
       - **Forward and Backward Propagation** to compute cost and gradients.
       - **Parameter Updates** using gradient descent.
       - **Prediction** for test and train datasets.
   * **Evaluation**:
     + Calculates accuracy for training and testing data.

**Outputs:**

* **Training Process**:
  + Cost decreases iteratively, indicating the model learns effectively.
* **Accuracy**:
  + Training and test accuracies are printed.

**Suggestions/Improvements:**

1. **Scaling and Normalization**:
   * Instead of manual min-max scaling, use sklearn.preprocessing.MinMaxScaler for better clarity and consistency.
2. **Learning Rate and Iterations**:
   * Tune these hyperparameters for optimal performance, especially if the dataset changes.
3. **Comparison with Libraries**:
   * Compare the custom implementation's performance with LogisticRegression from sklearn for verification.
4. **Model Validation**:
   * Perform **k-fold cross-validation** to assess model generalization.
5. **Feature Importance**:
   * Analyze feature importance to understand which attributes contribute most to predictions.
6. **Visualization**:
   * Add plots to compare predicted vs actual values for better insights.