

# Logistic Regression with Mini-Batch SGD Report

## Methodology

- 1. Data Loading & Preprocessing:**
  - Load the Wisconsin Breast Cancer dataset from Scikit-learn.
  - Split the dataset into training, validation, and test sets (70%-15%-15%).
  - Normalize features using `StandardScaler` to improve model convergence.
- 2. Model Training:**
  - Initialize weights from a standard Gaussian distribution.
  - Train using mini-batch SGD with a batch size of 64 and a fixed learning rate of 0.01.
  - Use binary cross-entropy as the loss function.
- 3. Evaluation:**
  - Compute predictions using the sigmoid function.
  - Evaluate model performance using accuracy, precision, recall, and F1-score.

## Results

- **Accuracy:** 94.18%
- **Precision:** 96.23%
- **Recall:** 94.44%
- **F1 Score:** 95.33%

## Conclusion

The logistic regression model with mini-batch SGD performs well on the Wisconsin Breast Cancer dataset, achieving high accuracy and balanced precision-recall metrics. Feature normalization significantly improves convergence and prevents numerical instability.

## Interpretation & Summary of Findings

The high accuracy and strong F1-score indicate that the model is effective at distinguishing between malignant and benign tumors. The precision score of 96.23% suggests that when the model predicts a tumor as malignant, it is correct most of the time. The recall of 94.44% shows that the model is also successful in identifying the majority of actual malignant cases. The balanced trade-off between precision and recall, reflected in the F1-score, makes this model suitable for medical diagnosis applications where both false positives and false negatives carry significant consequences. Further improvements could include hyperparameter tuning, regularization techniques, or experimenting with different optimization algorithms.