



Assignment 2: Machine Learning Foundations – Advanced Linear & Logistic Regression

Course: BCA – 3rd Year

Subject: Machine Learning Foundations

Topic: Advanced Linear and Logistic Regression

Submission Deadline: (Insert Date)



Objective

To enhance understanding of Linear and Logistic Regression by applying them to real-world datasets using feature scaling, regularization, and multi-feature modeling.



Part A: Multi-Feature Linear Regression

Create a Jupyter Notebook named `regression_models_assignment2.ipynb`.

Tasks:

1. Import Required Libraries
 - pandas, numpy, matplotlib, seaborn
 - sklearn.linear_model, sklearn.model_selection, sklearn.metrics, sklearn.preprocessing

2. Load Dataset

- Use a dataset like `house_prices.csv`, `insurance.csv`, or any with multiple numeric features

3. Data Exploration & Preprocessing

- Check for nulls, use `.describe()` and `.info()`
- Encode categorical features using `pd.get_dummies()`
- Use `MinMaxScaler` or `StandardScaler` on numeric features

4. Train-Test Split

- Use 75% for training and 25% for testing

5. Apply Linear Regression

- Train a `LinearRegression()` model on the dataset
- Evaluate with:
 - `r2_score`, `mean_absolute_error`, `mean_squared_error`

6. Plot Prediction vs Actual

- Use scatter plot or line chart to compare predictions with actual values



Part B: Logistic Regression with Regularization

Use a classification dataset such as:

- `breast_cancer` (from `sklearn.datasets`)
- Or `heart_disease.csv`, `diabetes.csv`

✓ Tasks:

7. Load & Preprocess Data

- Load dataset using `sklearn.datasets.load_breast_cancer()` or `pandas.read_csv()`
 - Handle missing values and normalize numeric features
8. Train-Test Split
- Split data into training and testing (80/20 split)
9. Train Logistic Regression with Regularization
- Use `LogisticRegression(penalty='l2', C=1.0)`
 - Train and predict using test data
10. Model Evaluation
- Display:
 - Confusion Matrix
 - Accuracy Score
 - Classification Report (Precision, Recall, F1)
 - Visualize Confusion Matrix using heatmap
11. Bonus (Optional)
- Plot ROC curve using `sklearn.metrics.roc_curve` and `auc`
-



Deliverables

- Jupyter Notebook: `regression_models_assignment2.ipynb`
 - Dataset (CSV or from sklearn)
 - Submit all in a folder named: `ML_Regression_Assignment2_YourName`
-



Evaluation Criteria

Criteria	Marks
Multi-feature Linear Regression	10
Logistic Regression with Regularization	10
Evaluation, graphs, and model interpretation	5
Clean code and structure	5
Total	30



Reference Resources

- [Scikit-learn – Linear Regression](#)
 - [Scikit-learn – Logistic Regression](#)
 - [ROC and AUC – sklearn docs](#)
 - [Feature Scaling – sklearn.preprocessing](#)
-