Logistic Regression

1. Introduction to Logistic Regression

Logistic Regression is a statistical method used for binary classification problems. Despite its name, it's a classification algorithm rather than a regression algorithm, used to predict the probability of a categorical dependent variable.

2. Types of Logistic Regression

2.1 Binary Logistic Regression

- Two possible outcomes (0 or 1)
- Examples:
    - Spam vs. Non-spam
    - Pass vs. Fail
    - Yes vs. No

2.2 Multinomial Logistic Regression

- Multiple categories (unordered)
- Examples:
    - Type of flower
    - Movie genre
    - Product category

2.3 Ordinal Logistic Regression

- Multiple ordered categories
- Examples:
    - Rating scales (1-5)
    - Education levels
    - Risk levels

3. Model Evaluation Metrics

3.1 Primary Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- ROC Curve
- AUC Score

3.2 Additional Metrics

- Confusion Matrix
- Log Loss
- McFadden's R-squared
- Classification Report

4. Implementation Steps

4.1 Data Preparation

1. Data cleaning
2. Feature selection
3. Feature scaling
4. Handling imbalanced data
5. Train-test split

4.2 Model Building

1. Feature engineering
2. Model initialization
3. Training
4. Validation
5. Hyperparameter tuning

4.3 Model Evaluation

1. Performance metrics
2. Cross-validation
3. Error analysis
4. Model comparison
5. Threshold optimization