# 1. Introduction to Logistic Regression

## What is Logistic Regression?

- **Logistic regression is a statistical method for analyzing datasets in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).**
- **Unlike linear regression, which predicts continuous outcomes, logistic regression predicts the probability of a binary outcome.**

## Comparison with Linear Regression

- **Linear Regression: Models a linear relationship between the input variables (X) and the output (Y), where Y is a continuous variable.**
- **Logistic Regression: Models the probability that a given input point belongs to a certain class. The output is a probability score between 0 and 1, which is then thresholded to classify the input.**

## Real-World Applications

- **Medical Diagnosis: Predicting whether a patient has a certain disease (yes/no).**
- **Credit Scoring: Assessing the likelihood of a borrower defaulting on a loan.**
- **Marketing: Predicting whether a customer will buy a product (purchase/no purchase).**

---

# 2. The Mathematical Foundation of Logistic Regression

## The Logistic Function (Sigmoid)

- **The logistic function, also known as the sigmoid function, is defined as:$\sigma(z)=11+e-z\sigma(z)=1+e-z1$**
- **This function maps any real-valued number into a value between 0 and 1, making it ideal for predicting probabilities.**

## Model Representation

- **The logistic regression model estimates the probability that a given input belongs to a certain class using the**

formula:P(Y=1∣X)=σ(β0+β1X1+β2X2+…+βnXn)P(Y=1∣X)=σ(β0+β1X1+β2X2+…+βnXn)

- Here, β0,β1,…,βnβ0,β1,…,βn are the parameters of the model.

### Probability Estimation

- The model output is a probability that is then used to predict the class label by applying a threshold (e.g., 0.5).

### Odds and Log-Odds

- Odds: The ratio of the probability of the event occurring to the probability of it not occurring:Odds=P(Y=1)1−P(Y=1)Odds=1−P(Y=1)P(Y=1)
- Log-Odds: The natural logarithm of the odds, which is modeled as a linear combination of the input variables:Log-Odds=log(P(Y=1)1−P(Y=1))=β0+β1X1+…+βnXnLog-Odds=log(1−P(Y=1)P(Y=1))=β0+β1X1+…+βnXn

---

## 3. Training a Logistic Regression Model

### Maximum Likelihood Estimation (MLE)

- MLE is used to estimate the parameters (ββ) of the logistic regression model by maximizing the likelihood function:L(β)=∏i=1nP(Yi∣Xi)L(β)=i=1∏nP(Yi∣Xi)
- The goal is to find the parameter values that make the observed data most probable.

### Optimization Techniques

- Gradient Descent: An iterative optimization algorithm used to minimize the loss function (negative log-likelihood) by adjusting the model parameters.
- Variants of gradient descent include Stochastic Gradient Descent (SGD) and Mini-batch Gradient Descent.

### Regularization

- Regularization techniques like L1 (Lasso) and L2 (Ridge) are used to prevent overfitting by penalizing large coefficients.
- L1 Regularization: Encourages sparsity in the model, potentially setting some coefficients to zero.
- L2 Regularization: Encourages small coefficients but doesn't enforce sparsity.

---

# 4. Model Evaluation and Metrics

## Confusion Matrix

- **The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known.**
- **Components:**
    - **True Positives (TP), False Positives (FP)**
    - **True Negatives (TN), False Negatives (FN)**

## Classification Metrics

- **Accuracy: The ratio of correctly predicted observations to the total observations.Accuracy=TP+TNTP+TN+FP+FNAccuracy=TP+TN+FP+FNTP+TN**
- **Precision: The ratio of correctly predicted positive observations to the total predicted positives.Precision=TPTP+FPPrecision=TP+FPTP**
- **Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in the actual class.Recall=TPTP+FNRecall=TP+FNTP**
- **F1 Score: The harmonic mean of precision and recall.F1=2×Precision×RecallPrecision+RecallF1=2×Precision+RecallPrecision×Recall**
- **ROC Curve: Plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.**
- **AUC (Area Under the Curve): Measures the entire two-dimensional area underneath the ROC curve. Higher AUC indicates better model performance.**

## Calibration of Probabilities

- **Calibration ensures that predicted probabilities match the actual probabilities. Techniques include:**
    - **Platt Scaling: Fits a logistic regression model to the classifier's scores.**
    - **Isotonic Regression: A non-parametric method that fits a piecewise-constant non-decreasing function.**

---

# 5. Multinomial and Ordinal Logistic Regression

## Multinomial Logistic Regression

- **Used when the dependent variable is categorical with more than two levels (multi-class classification).**
- **The model generalizes logistic regression using the softmax function, which outputs a probability distribution over multiple classes.**

### Ordinal Logistic Regression

- **Applied when the dependent variable is ordinal, meaning the categories have a meaningful order but the distances between them are not known.**
- **The model predicts the probability of each possible outcome using cumulative logits.**

### Use Cases

- **Multinomial: Predicting the type of cuisine (Italian, Mexican, Chinese) a restaurant serves based on its features.**
- **Ordinal: Predicting a satisfaction level (unsatisfied, neutral, satisfied) from survey data.**

---

## 6. Assumptions and Limitations

### Linearity in Log-Odds

- **The model assumes a linear relationship between the independent variables and the log-odds of the dependent variable.**

### Independence of Observations

- **Each observation is assumed to be independent of others, which may not hold in cases of time series or grouped data.**

### Multicollinearity

- **Occurs when independent variables are highly correlated, which can make it difficult to determine the effect of each variable. Solutions include removing or combining correlated variables.**

### Handling Non-Linearity

- **Logistic regression struggles with non-linear relationships between independent and dependent variables. Non-linear transformations or polynomial features can be used to address this.**

---

# 7. Advanced Topics in Logistic Regression

## Regularization Paths

- **Regularization paths visualize how the coefficients of a logistic regression model change as the regularization parameter (e.g., lambda) varies.**
- **This helps in understanding the impact of regularization and selecting the optimal value of the regularization parameter.**

## Penalized Logistic Regression

- **In penalized logistic regression, penalty terms are added to the loss function to control the complexity of the model, thereby preventing overfitting.**

## Feature Selection

- **Logistic regression can be used for feature selection by evaluating the significance of each feature's coefficient. Regularization techniques like Lasso also help in automatically selecting relevant features by setting some coefficients to zero.**

## Logistic Regression with Interaction Terms

- **Interaction terms allow modeling interactions between variables, where the effect of one variable depends on the value of another. The interpretation of coefficients becomes more complex but can reveal important insights.**

---

# 8. Practical Implementation in Python

## Logistic Regression in Python

- **Libraries: Scikit-learn provides an easy-to-use interface for implementing logistic regression.**
- **Steps:**
    1. **Data Preprocessing: Handling missing values, encoding categorical variables, and feature scaling.**
    2. **Model Training: Using `LogisticRegression` from Scikit-learn.**
    3. **Model Evaluation: Evaluating the model using the confusion matrix, classification metrics, and ROC curve.**

4. **Prediction: Making predictions on new data.**

### Interpreting the Output

- **Coefficients: Represent the change in the log-odds of the outcome for a one-unit increase in the predictor variable.**
- **P-values: Indicate the significance of each predictor variable in the model.**
- **Odds Ratios: Exponentiating the coefficients gives the odds ratios, which are easier to interpret in practical terms.**

---

## 9. Ethical Considerations and Best Practices

### Bias and Fairness

- **Logistic regression models can inherit biases from the training data, leading to unfair predictions. Techniques like fairness constraints or re-weighting can be used to mitigate bias.**
-