Difference Between Logistic Regression and Naive Bayes (Simplified)

1. Purpose:

Both Logistic Regression and Naive Bayes are used for Classification tasks.

2. How They Approach Probability:

Logistic Regression:

- Directly models the probability of a class given the features.
- Uses the sigmoid function to produce probabilities.
- Example: "What is the probability this email is spam?"

Naive Bayes:

- Uses Bayes Theorem to flip the problem.
- Calculates how likely the features are given a class.
- Example: "How likely are these words to appear in spam emails?"

3. Feature Assumptions:

Logistic Regression:

- Assumes a linear relationship between features and the log-odds of the outcome.
- Does NOT assume features are independent.

Naive Bayes:

- Assumes all features are independent of each other given the class (naive assumption).
- Simplifies calculations but is often not true in reality.

4. Model Type:

Logistic Regression:

- A discriminative model Focuses on finding the boundary between classes.

Naive Bayes:

- A generative model Focuses on modeling how the data was generated.

5. Use Cases:

Logistic Regression:

- Works well when features interact or are correlated.
- Often better with numerical data.

Naive Bayes:

- Works well for text data (e.g., spam detection, sentiment analysis).
- Fast and performs well even with small datasets.

Summary Table:

Aspect	Logistic Regression	Naive Bayes
Type	Discriminative	Generative
Probability Approach	Models P(Class | Features)	Models P(Features | Class)
Feature Assumptions	No independence assumed	Assumes feature independence
Typical Data	Numeric, mixed	Text, categorical
Speed	Slower to train	Very fast