Lecture: Logistic Regression — Calculating Probabilities with the Sigmoid Function

1. Introduction

Many real-world machine learning problems require predicting **probabilities**, not just raw numerical values. Logistic regression is a widely used and computationally efficient model designed specifically for this purpose.

Unlike linear regression, which predicts any real-valued number, logistic regression produces outputs strictly between **0 and 1**, allowing the output to be interpreted as a **probability**.

---

2. How Logistic Regression Outputs Are Used

The probability produced by logistic regression can be used in two main ways:

1. **Used directly as a probability**
   Example:
   If a model outputs 0.932, this indicates a 93.2% probability that an email is spam.

2. **Converted into a binary class**
   Example:

   - Probability ≥ 0.5 → Spam
   - Probability < 0.5 → Not Spam

This lecture focuses on the **probability output itself**. Binary classification decisions are covered separately.

---

3. The Sigmoid (Logistic) Function

Definition

The sigmoid function is the mathematical function that enables logistic regression to output valid probabilities.

σ(x)=11+e−x\sigma(x) = \frac{1}{1 + e^{-x}}σ(x)=1+e−x1

Where:

- σ(x)\sigma(x)σ(x) is the output of the function
- eee is Euler's number (≈ 2.71828)
- xxx is the input value

Key Properties of the Sigmoid Function

- Output is always between **0 and 1**
- Output approaches 1 as x→+∞x \to +\inftyx→+∞
- Output approaches 0 as x→−∞x \to -\inftyx→−∞
- Output is exactly 0.5 when x=0x = 0x=0
- The curve is **S-shaped** (sigmoid)

These properties make the sigmoid function ideal for probability modeling.

---

4. Linear Component of Logistic Regression

Before applying the sigmoid function, logistic regression computes a **linear combination of features**, similar to linear regression.

z=b+w1x1+w2x2+⋯+wnxnz = b + w_1x_1 + w_2x_2 + \dots + w_nx_nz=b+w1x1+w2x2+⋯+wnxn

Where:

- zzz is the **linear output** (also called *log-odds*)
- bbb is the bias
- wiw_iwi are the learned weights
- xix_ixi are the feature values

This linear output can take **any real value**, positive or negative.

---

5. Transforming Linear Output into a Probability

To convert the linear output into a probability, logistic regression applies the sigmoid function:

y^=σ(z)=11+e−z\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}y^=σ(z)=1+e−z1

Where:

- y^\hat{y}y^ is the predicted probability

This transformation ensures:

- Extreme negative values map close to 0
- Extreme positive values map close to 1

---

6. Visual Intuition

- The linear function produces a straight line that extends infinitely in both directions
- The sigmoid function **bends** this line into an S-shaped curve
- Large negative inputs produce probabilities near 0
- Large positive inputs produce probabilities near 1

Example mappings:

- z=−10⇒y^≈0.00004z = -10 \Rightarrow \hat{y} \approx 0.00004z=−10⇒y^≈0.00004
- z=0⇒y^=0.5z = 0 \Rightarrow \hat{y} = 0.5z=0⇒y^=0.5
- z=5⇒y^≈0.9933z = 5 \Rightarrow \hat{y} \approx 0.9933z=5⇒y^≈0.9933

---

7. Example: Calculating Logistic Regression Output

Given:

- A logistic regression model with three features
- Known bias and weights
- Known input feature values

Step 1: Compute the Linear Output zzz

z=b+w1x1+w2x2+w3x3z = b + w_1x_1 + w_2x_2 + w_3x_3z=b+w1x1+w2x2+w3x3

Substituting the given values yields:

z=1z = 1z=1

Step 2: Apply the Sigmoid Function

y^=11+e−1\hat{y} = \frac{1}{1 + e^{-1}}y^=1+e−11 y^≈0.731\hat{y} \approx 0.731y^≈0.731

---

8. Interpretation of the Result

- The linear output z=1z = 1z=1 indicates positive log-odds
- The sigmoid-transformed output of **0.731** means:
    - There is a **73.1% probability** of the positive class

---

9. Key Concepts: Log-Odds

- The linear output zzz represents **log-odds**
- Log-odds measure confidence before probability transformation
- Positive log-odds → probability greater than 0.5
- Negative log-odds → probability less than 0.5

---

10. Summary

- Logistic regression is designed to output probabilities
- The sigmoid function ensures outputs lie strictly between 0 and 1
- Logistic regression first computes a linear combination of features
- The sigmoid function converts this linear output into a probability
- The output can be interpreted directly or used for classification

---

11. Key Terms

- **Logistic regression**: A probabilistic classification model
- **Sigmoid function**: Function that maps real numbers to (0, 1)
- **Log-odds (z)**: Linear output before sigmoid transformation
- **Probability output**: Final prediction of logistic regression