Logistic Regression

It is not possible to solve a classification problem with simple linear regression. Linear regression establishes a linear relationship between features and output and predicts continuous values. Since linear regression maps the output to a straight line, where the output could be negative as well as more than 1, it is not possible to classify features into discrete classes.

Logistic Regression solves this problem by multiplying the output of a **linear function** with a **sigmoid function**, to map the output to a range of 0 to 1.

This digital note is part of my machine learning docs; find more at my GitHub profile.

Mathematical Intuition

Mathematically, we multiply the output of a linear function (like linear regression) with a sigmoid function.

For a binary classification problem, the linear function could be as simple as multiplying the values of x by some specified weights:

$$z = w_1x_1 + w_2x_2 + \ldots + w_nx_n + b$$

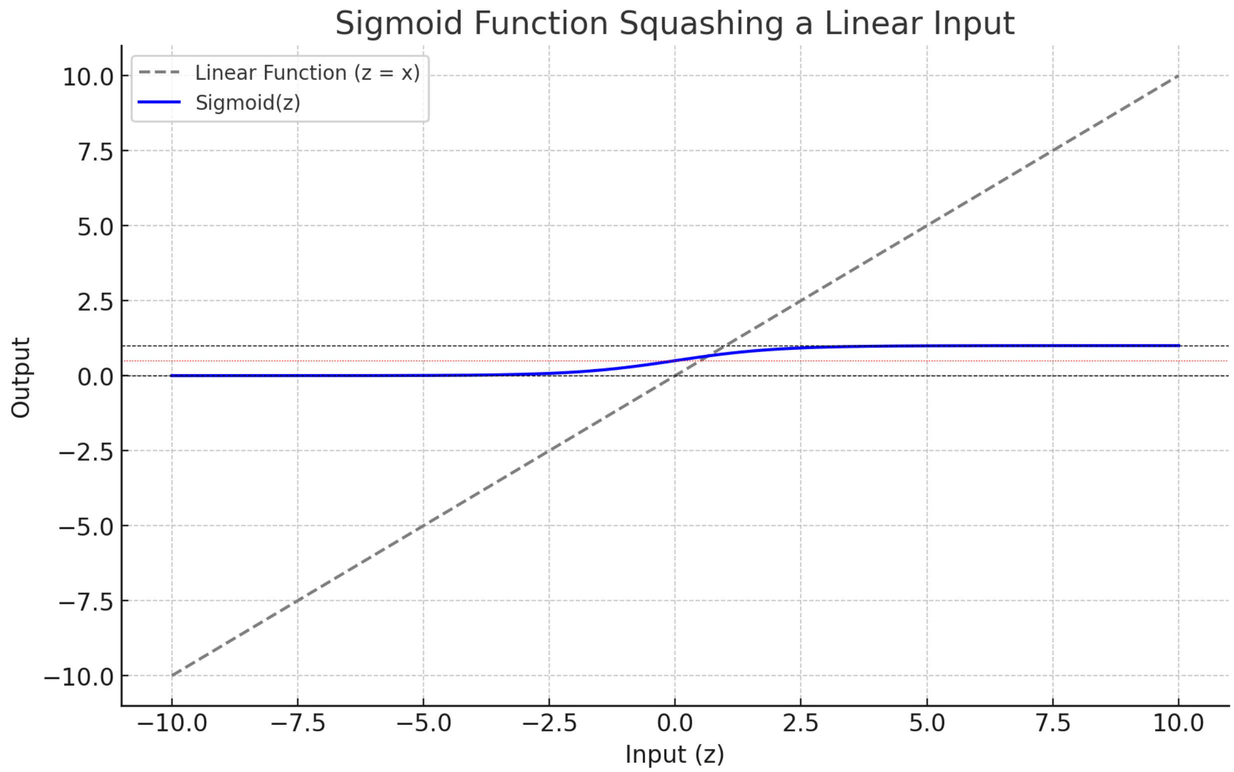
The output of this function is multiplied by sigmoid function to map it into a range of 0 to 1.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- **e** is Euler's number (2.71828)
- **x** is the output of the previously used linear function.
- x ≥ 0.5, output is predicted to be class 1.
- x<0.5, output is predicted to be class 0.
- x will always be in a range of 0 to 1 (it is never exactly 0 or 1)

Geometric Intuition

Geometrically, we aim to squash the straight line produced by a linear function using sigmoid.

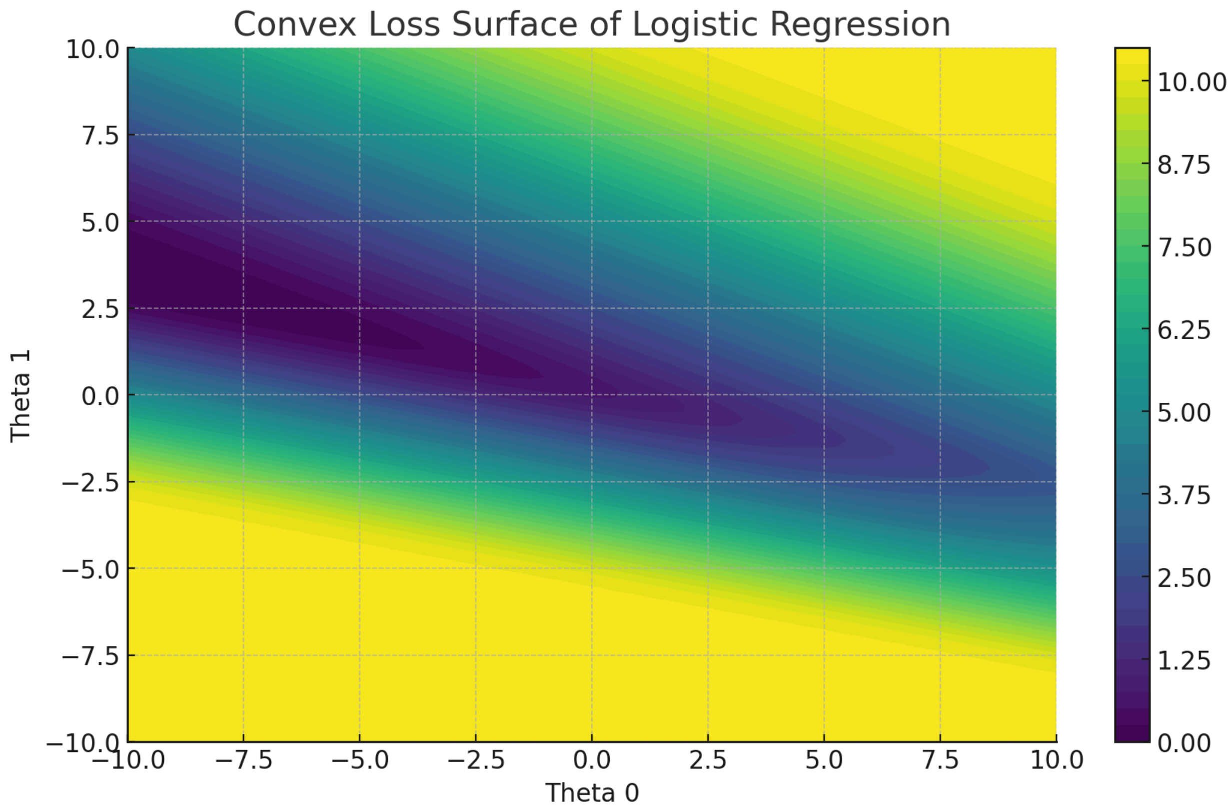


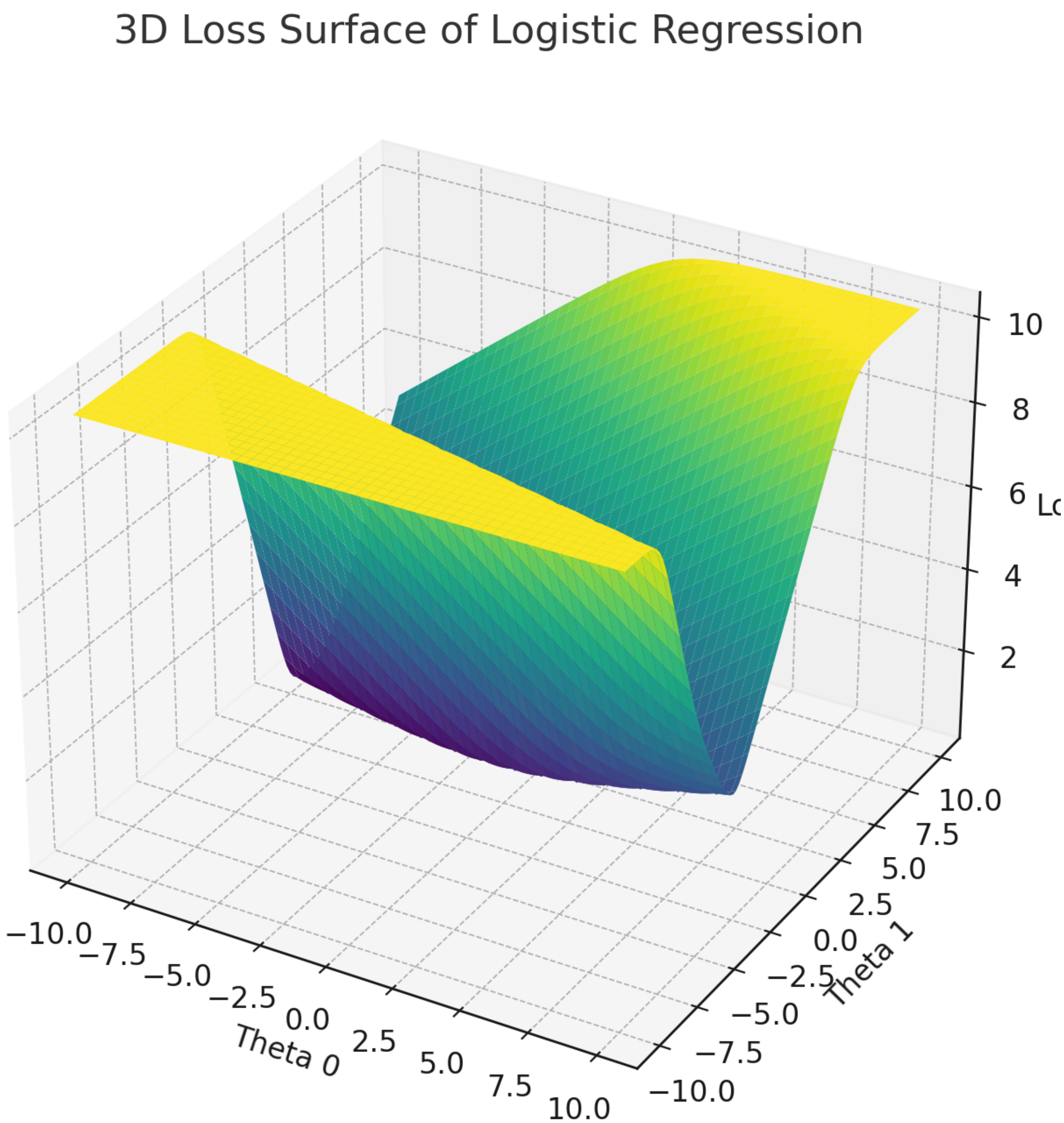
Loss Function

The error for logistic regression can be calculated with **Binary Cross-Entropy Loss** (**Log Loss**).

$$\mathcal{L}(y, \hat{y}) = -\left(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right)$$

Even though we use the sigmoid function, the loss function of logistic regression will always produce a **convex gradient**, which guarantees **convergence to global minima** when an appropriate **learning rate** is chosen.


Convex Loss Surface of Logistic Regression
Theta 1
Theta 0
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
10.00
8.75
7.50
6.25
5.00
3.75
2.50
1.25
0.00


3D Loss Surface of Logistic Regression
Theta 0
Theta 1
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
2
4
6
8
10