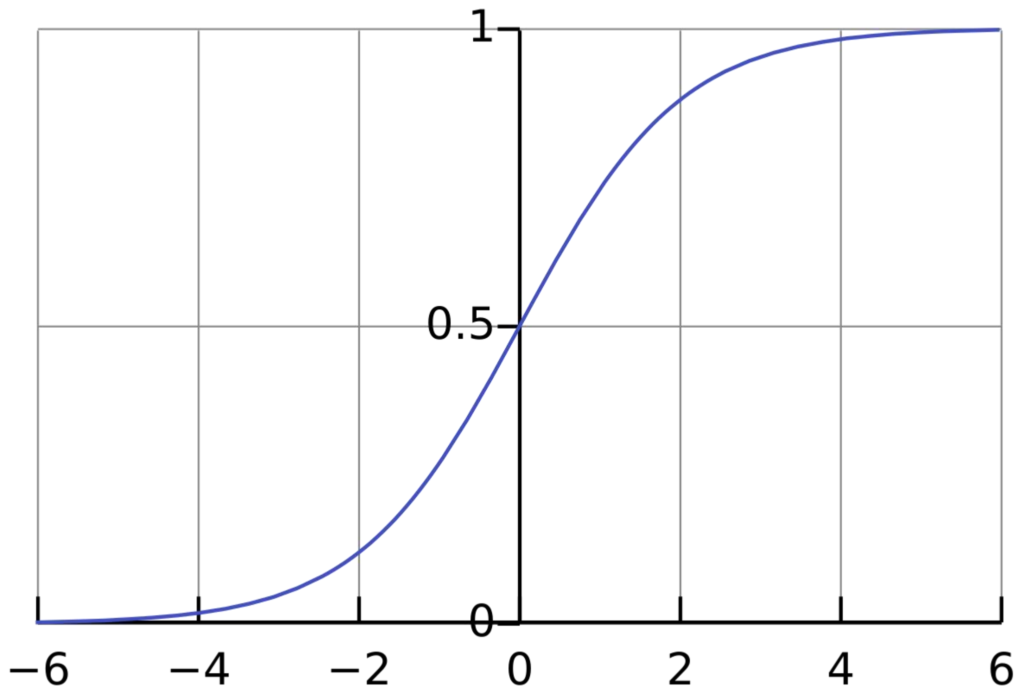
Logistic Regression

Theo như mình đã biết được từ Linear Regression, mình sẽ đi tìm đường thẳng tuyến tính phù hợp nhất cho các điểm dữ liệu trong một tập của mình với công thức:

$$\hat{y} = wx + b$$

Nhưng đối với Logistic Regression (Hồi quy Logistic), mình đang cần tìm các xác suất hay vì một giá trị cụ thể. Và để làm được việc đó, mình sẽ đặt giá trị của mình vào hàm Sigmoid.

Hàm Sigmoid



Hàm này cho phép mình lấy về một phân phối xác suất nằm giữa 0 và 1. Nó có công thức sau:

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

Và công thức tính kết quả dự đoán/đầu ra của thuật toán Logistic Regression là:

$$\hat{y} = h_0(x) = \frac{1}{1 + e^{-(wx + b)}}$$

Khác với Linear Regression, mình sử dụng Cross-Entropy làm hàm mất mát của thuật toán.

$$J(w,b) = \frac{1}{n} * \sum_{i=1}^{n} [y_i * \log(\hat{y}) + (1 - y_i) * \log(1 - \hat{y})]$$

Và để sử dụng thuật toán Gradient Descent, mình cần tính độ dốc của hàm mất mát lần lượt cho trọng số và bias.

$$J^{'}(w,b) = \begin{bmatrix} \frac{dJ}{dw} \\ \frac{dJ}{db} \end{bmatrix} = \begin{bmatrix} \frac{-1}{n} * \sum x_i(y_i - \hat{y}) \\ \frac{-1}{n} * \sum (y_i - \hat{y}) \end{bmatrix}$$

Gradient Descent cũng được tính tương tự như bên Linear Regression để cập nhật trọng số và bias qua từng vòng lặp.

$$w = w - \alpha * \frac{dJ}{dw}$$

$$b = b - \alpha * \frac{dJ}{db}$$

Thuật toán sẽ cập nhật trọng số và bias cho đến khi nào mà hàm mất mát đạt đến giá trị nhỏ nhất. Rồi mình sẽ dùng chúng để dự đoán các đầu vào mới để có kết quả. Kết quả này thường là một xác suất.

Sự khác biệt giữa Linear Regression và Logistic Regression là:

- Linear Regression dùng để dự đoán các giá trị số liên tục.
- Logistic Regression là một dạng phân loại nhị phân, tức nó chỉ có thể phân loại giữa 2 biến số 0 và 1. Ví dụ: không và có; nam và nữ; …

Kết quả của thuật toán Logistic Regression thường là một xác suất nằm giữa con số 0 và 1. Và kết quả phân loại thường dựa vào hàm sigmoid. Nếu kết quả dự đoán xác suất trên 0.5, thì phân loại bằng 1. Còn lại bằng 0.