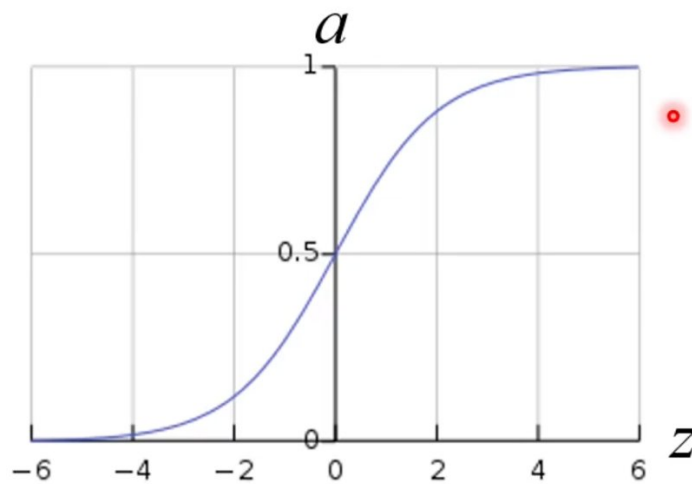


MỘT SỐ ACTIVATION PHỔ BIẾN

1. Sigmoid

$$G(z) = \frac{1}{1 + e^{-z}}$$



*Ưu điểm:

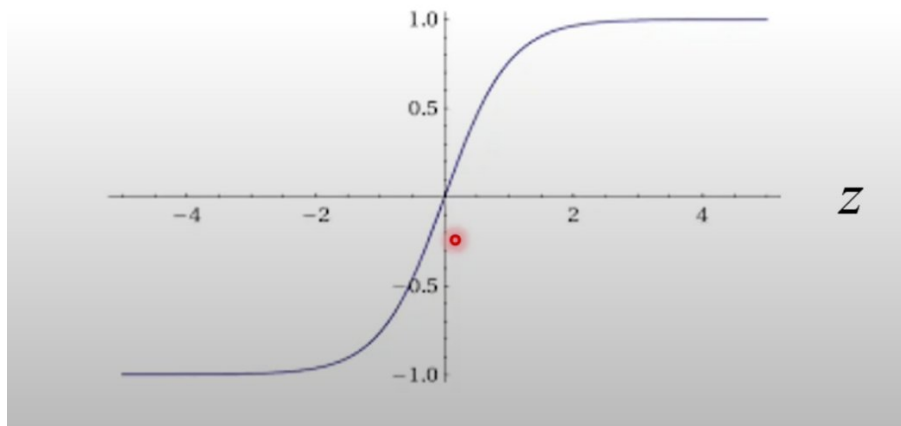
+ Output đầu ra luôn nằm trong đoạn $[0, 1]$, phù hợp trong vấn đề phân loại binary 0,1, dễ dàng tính gradient.

*Nhược điểm:

+ Nếu z nhỏ thì ta nhận được output vô cùng bé có thể về 0. Nó sẽ làm mất đi đặc tính của output.

2. Hyperbolic Tangent (tanh) function

$$G(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



* Ưu điểm:

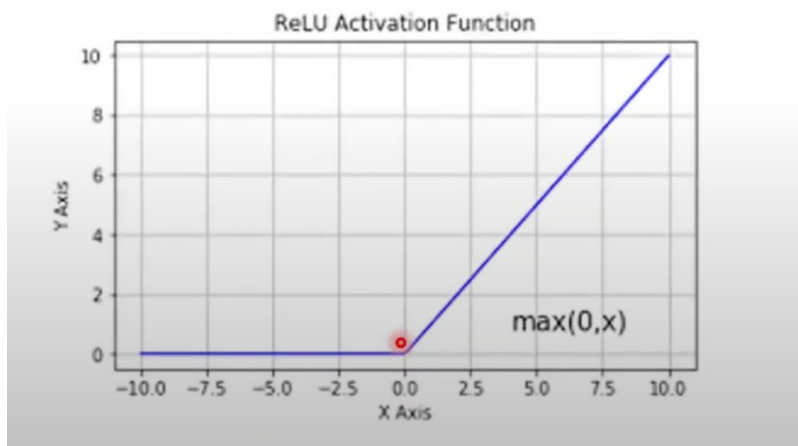
+ Giá trị output nằm trong khoảng từ -1 đến 1. Tốt hơn mô hình sigmoid ở chỗ đầu ra bằng 0 khi $z=0$.

* Nhược điểm:

+ Tuy vậy nó vẫn phần nào mất đi đặc tính của trọng số. Và độ phức tạp tính toán lớn.

3. Rectified Linear Unit (ReLU)

$$G(z) = \max(0, z)$$



*Ưu điểm:

+Tính toán nhanh, gradient của giá trị đầu vào dương vẫn được giữ nguyên. Thường được sử dụng phổ biến hơn

*Nhược điểm:

+Tuy vậy, nếu đầu vào nhỏ hơn 0 thì sẽ quy hết về không. Chính vì vậy để khắc phục nhược điểm này ta có hàm.

Leaky ReLU

$$\sigma(z) = \begin{cases} \alpha z, & \text{if } z < 0 \\ z, & \text{if } z \geq 0 \end{cases}$$

4. Softmax function

$$g(z)_i = \frac{e^{z_i}}{\sum_{i=1}^N e^{z_i}}$$

*Ưu điểm:

+Tổng của tất cả các lớp bằng 1, điều này giúp dễ dàng diễn giải đầu ra của mô hình, mỗi đầu ra có thể hiểu là xác suất dự đoán của từng lớp.

*Nhược điểm:

+ Dễ bị ảnh hưởng bởi các lớp có xác suất cao