

Swish激活

Swish激活函数是现代神经网络中的一种常用非线性激活函数，其公式如下：

$$\text{Swish}(x) = x \cdot \sigma(x) \quad (1)$$

其中， $\sigma(x)$ 是标准的 Sigmoid 函数，定义为：

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

Swish激活的特点

1. 平滑性：

- Swish 是连续且可微的，与传统的 ReLU 激活不同，Swish 在 $x = 0$ 附近不会产生梯度突然变化。

2. 自调节性：

- Swish 的值会根据输入 x 的大小动态调节。在 $x > 0$ 时，Swish 函数接近恒等映射 $f(x) \approx x$ ；而在 $x < 0$ 时，Swish 函数接近 0，但并不会完全为 0。

3. 负值处理：

- Swish 的负输出在一定程度上保留了负输入的信息，而不像 ReLU 那样直接截断。

4. 实验性能：

- Swish 激活在许多任务上表现优于 ReLU，尤其是在现代深度学习模型（如 EfficientNet）中。

Swish的分布特性

Swish 的负值范围较窄，其值大致分布在 $[-0.278, 0]$ ，而正值部分可以是无界的。这种不对称性使得标准对称或无符号量化方法难以适应，从而导致信息丢失。

在文章中，Swish 激活函数的这些特性对量化提出了以下挑战：

• 负值范围处理：

- 标准的无符号量化会将所有负值置为零，这会丢失 Swish 激活负值部分的重要信息。
- 对称量化虽然保留了负值，但未充分利用正值较宽范围的动态信息。

为此，**LSQ+** 引入了可学习的偏移量（**offset**），实现非对称量化，能够动态调整量化范围，使其更好地适应 Swish 的不对称分布，从而显著提升量化后的模型性能。