Self-Gated Activation Function

Presented by Yvan Tamdjo Biakeu

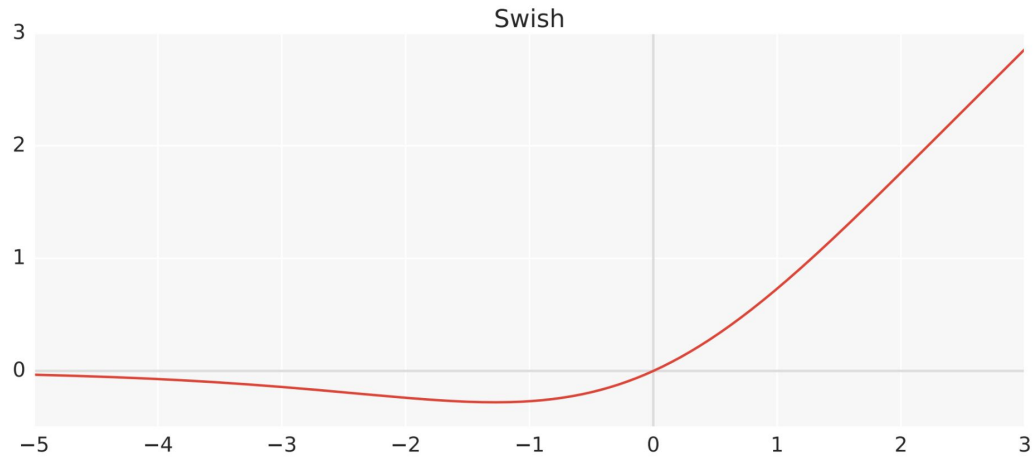
Core idea of the paper

- The effect of the activation on the training dynamics and task performance in deep network
- The most successful and widely-used activation function is Rectified Linear Unit(ReLU)
- Many alternatives of ReLU have been proposed
- The paper show another activation function called Swish that outperforms or matches ReLu on variety of Deep model

Swish

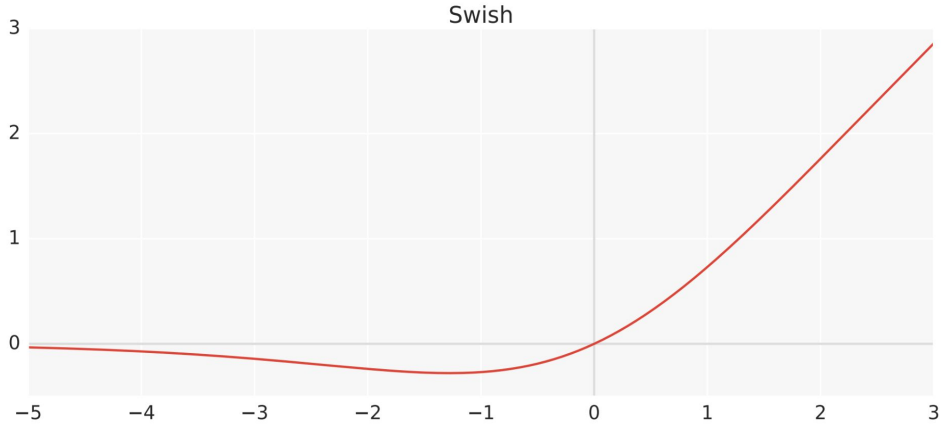
swish(x) = x * sigmoid(x)

Where **sigmoid(x) = 1 / (1 + exp(-x))**



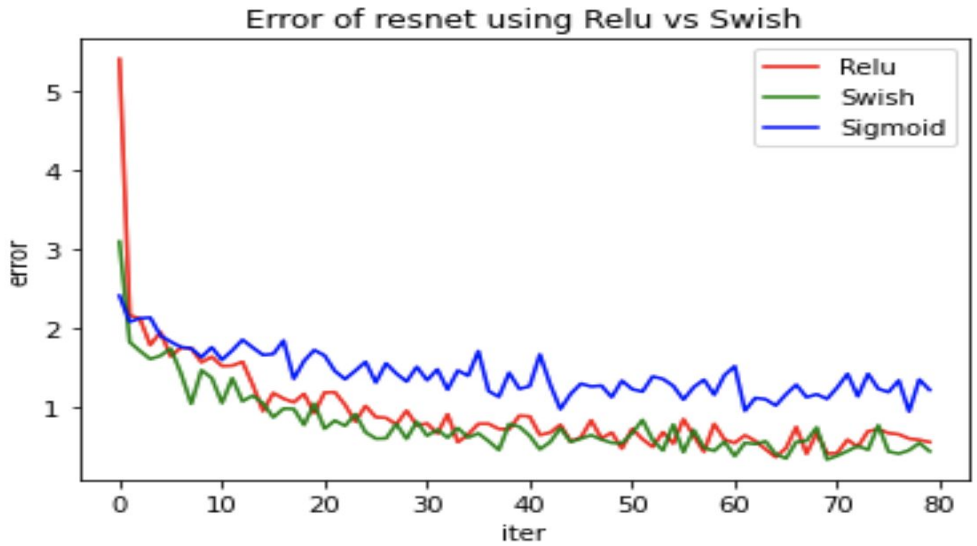
Properties of Swish

- Non-monotonic function
- Unbounded above and bounded below function
- Smooth function



Experiment

- We have done the experiment with a classification task on CIFAR10 dataset using Resnet32
- The result show that Swish outperforms or matches ReLU



Thank you