

Transformerの構成要素：正規化

- 正規化の方法

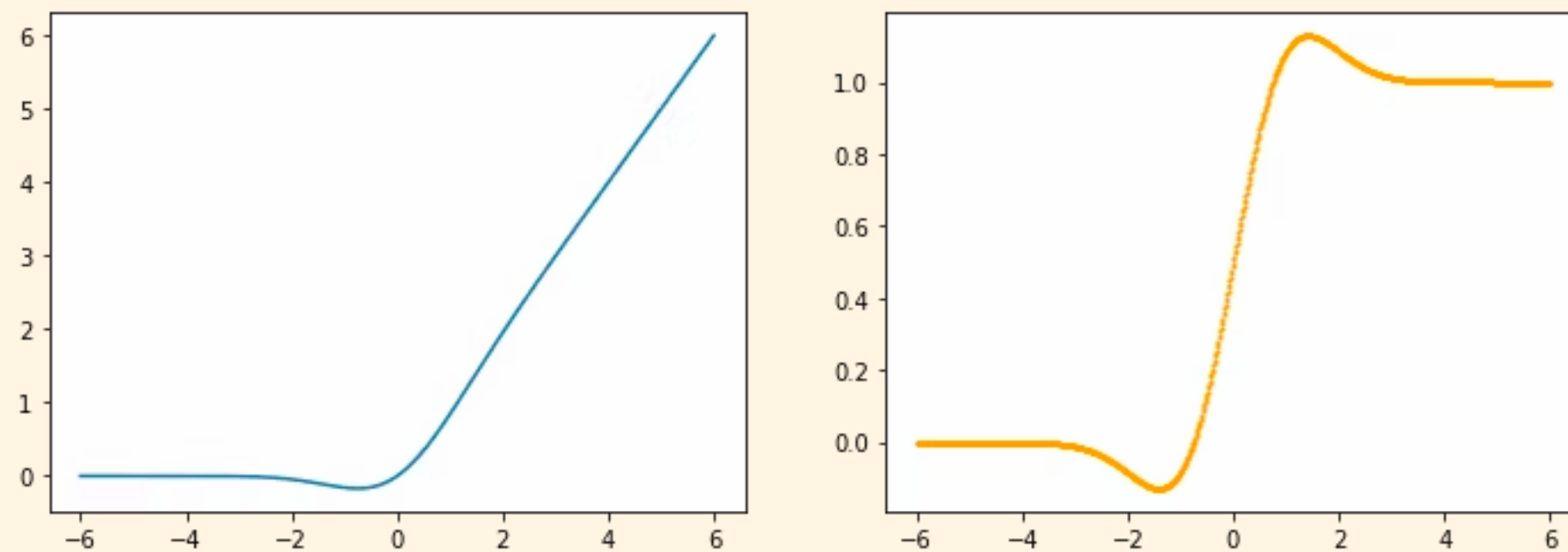
- ❖ 学習を安定させるため、正規化は広く採用されている
- ❖ 初期のTransformerはLayerNorm (平均と分散を用いて正規化する手法) を用いていたが、近年ではRMSNorm (二乗平均平方根を用いた手法) や DeepNorm (スキップ接続をスケールして学習を安定させる手法) といった手法も提案されている

- 正規化の位置

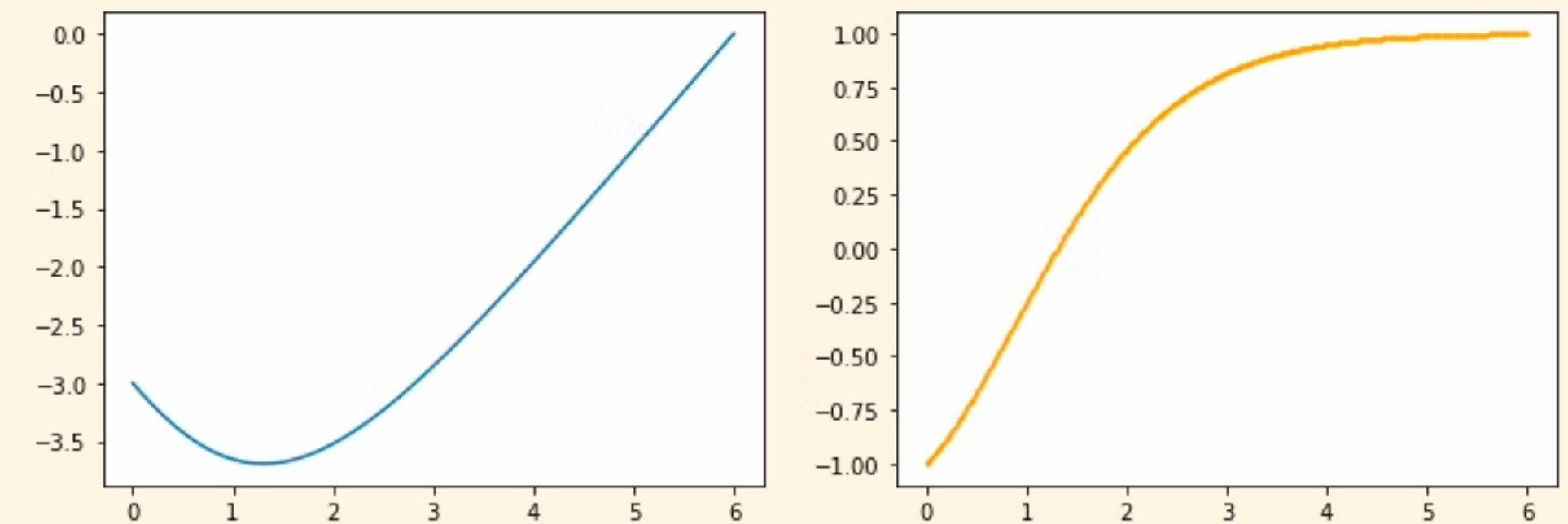
- ❖ 学習の安定性から、ほとんどのLLMで各サブレイヤーと最終予測の前に正規化を置く pre-LNが採用されている

Transformerの構成要素：活性化関数

- 活性化関数は、ニューラルネットワークの表現力を高めるために一般的に用いられる
- 既存のLLMでは、GeLU活性化関数が広く使われている
- 最新のLLM (PaLMやLaMDA) では、GLU活性化関数の系統も使用されており、特にSwiGLUとGeGLUの変種は良い性能を示す



GeLU



GLU