

[Novak et. al 2019] [Lee et. al 2018]

[Arora et. al 2019] [Cao and Gu, 2019] [Jacot et. al 2017]

(lazy training) CK/GPK

(full training) NTK

DNN-NTK > 5%

Infinite Width DNN = Kernels

Finite Width Deep Neural Networks

Duality: In DNNs with ReLU most information is in the gates

[Lakshminarayanan and Singh, 2020]

Simplifying NTK

Simple Experiments

$\text{NTK} = \text{NTK}^{\text{fixed-gate}} + \text{NTK}^{\text{gate-learning}}$
 $\text{NTK}^{\text{fixed-gate}} = \text{constant} \times \text{NPK}$

Deep Gated Networks (DGN)
DNN-DGN < 1%

[This Paper]

NPK is a composite kernel

[This Paper]

Generalisation With

Depth & Width Skip Convolution
+ Pooling

Layer Constant Gate Only
Permutations Input Training