Loss Surfaces, Mode Connectivity, and Fast Ensembling of DNNs

Timur Garipov\textsuperscript{1,2} Pavel Izmailov\textsuperscript{3} Dmitrii Podoprikhin\textsuperscript{4} 
Dmitry Vetrov\textsuperscript{5} Andrew Gordon Wilson\textsuperscript{3}

\textsuperscript{1}Samsung AI Center in Moscow, \textsuperscript{2}Skolkovo Institute of Science and Technology, \textsuperscript{3}Cornell University, \textsuperscript{4}Samsung-HSE Laboratory, \textsuperscript{5}National Research University Higher School of Economics

Neural Information Processing Systems
Montreal, Canada

December 4, 2018
Loss Surfaces

ResNet-164, CIFAR-100
Loss Surfaces

ResNet-164, CIFAR-100
Finding Paths between Modes

- Weights of pretrained networks: \( \hat{w}_1, \hat{w}_2 \in \mathbb{R}^{|\text{net}|} \)

- Define parametric curve: \( \phi_\theta(\cdot) [0, 1] \rightarrow \mathbb{R}^{|\text{net}|} \)

  \[
  \phi_\theta(0) = \hat{w}_1, \quad \phi_\theta(1) = \hat{w}_2
  \]

- DNN loss function: \( \mathcal{L}(w) \)

- Minimize averaged loss w.r.t. \( \theta \)

\[
\minimize{\theta} \int_0^1 \mathcal{L}(\phi_\theta(t))dt = \mathbb{E}_{t \sim U(0, 1)} \mathcal{L}(\phi_\theta(t))
\]
Loss Surfaces

VGG-16, CIFAR-10

Train loss

Test error (%)
Fast Geometric Ensembles (FGE)

Learning Rate

75% training

Ensemble

Epoch

Distance

α₁

α₂

n

Test error (%)

0 0.5 1 1.5 2 2.5 3 3.5

FGE iteration number

0 0.5c 1c 1.5c 2c 2.5c 3c 3.5c
Ensembling Results

SSE = Huang et al., (‘‘Snapshot ensembles: Train 1, get m for free’’), ICLR 2017
Summary

- Local optima are connected by simple curves.
- To find these curves we minimize loss uniformly in expectation over a path from one mode to another.
- We are inspired by these insights to propose a fast ensembling algorithm.

PyTorch code released for both mode connectivity and FGE

Come to our poster #162!