Identity Curvature Laplace Approximation for Improved Out-of-Distribution Detection

Maksim Zhdanov @ AIRI, NUST MISIS; Stanislav Dereka @ T-Bank, MIPT; Sergey Kolesnikov @ T-Bank

Laplace Approximation for OOD Detection?

In our paper, we approach the OOD detection task with Laplace approximation. We show the misalignment between the properties of the Fisher information matrix (FIM) used in Laplace approximation and the distribution of classes in common datasets, which limits the OOD detection performance. For instance, the long tail of FIM's spectral distribution does not help reflect a more uniform distribution of class embeddings for CIFAR-10 and CIFAR-100 datasets.

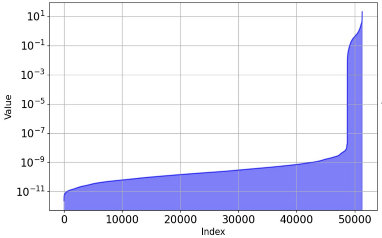

Figure 1. Spectral distribution of empirical Fisher

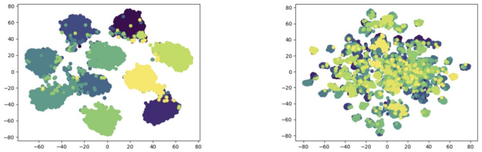

Figure 2. 2D projections of CIFAR-10 (left) and CIFAR-100 (right) model embeddings.

Identity Curvature Laplace Approximation

Such issues hinder the Laplace approximation performance on OOD detection. Thus, we present ICLA, a Laplace approximation without all second-order information. Our approach results in better OOD detection and lowers computation costs. We compare ICLA with common Laplace approximation variations and state-of-the-art non-Bayesian OOD detection methods and show its competitive performance.

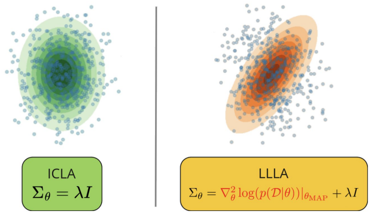

Figure 3. ICLA

Method	CIFAR-10		CIFAR-100		ImageNet-200	
	Near OOD	Far OOD	Near OOD	Far OOD	Near OOD	Far OOD
MSP	88.03 ± 0.25	90.73 ± 0.43	80.27 ± 0.11	77.76 ± 0.44	**83.34 ± 0.06**	90.13 ± 0.09
ODIN	82.87 ± 1.85	87.96 ± 0.61	79.90 ± 0.11	79.28 ± 0.21	80.27 ± 0.08	91.71 ± 0.19
ReAct	87.11 ± 0.61	90.42 ± 1.41	80.77 ± 0.05	80.39 ± 0.49	81.87 ± 0.98	92.31 ± 0.56
VIM	88.68 ± 0.28	**93.48 ± 0.24**	74.98 ± 0.13	81.70 ± 0.62	78.68 ± 0.24	91.26 ± 0.19
SHE	81.54 ± 0.51	85.32 ± 1.43	78.95 ± 0.18	76.92 ± 1.16	80.18 ± 0.25	89.81 ± 0.61
ASH	75.27 ± 1.88	78.49 ± 2.31	78.20 ± 2.21	80.58 ± 2.56	79.38 ± 1.93	**92.74 ± 1.91**
ICLA	90.01 ± 0.21	92.50 ± 0.38	**81.45 ± 0.10**	80.79 ± 0.46	81.88 ± 0.10	89.45 ± 0.14
ICLA$_{zero}$	**90.56 ± 0.23**	93.18 ± 0.45	81.38 ± 0.28	**82.49 ± 0.60**	80.70 ± 0.10	89.82 ± 0.11

Table 1. OOD detection results.

Why Laplace Approximation Might not Work in Your Case, but ICLA will

We analyze the root of this intriguing result, showing that the performance of classical Laplace approximation depends on the structure of the selected dataset. We demonstrate how the difference in performance between the classical Laplace approximation and ICLA vanishes when moving from a highly separable dataset to a less separable one. This experiment indicates that standard Laplace approximation fails for datasets with separable classes. ICLA does not have such an issue.

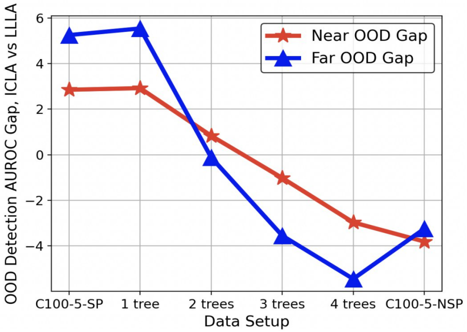

Figure 4. Data separability experiment. Gradually moving from a separable (C100-5-SP) to less separable dataset (C100-5-NSP), we observe how standard LLLA begins to estimate uncertainty better than its alternative with identity curvature.

zhdanov@airi.net