Your 598 ID: 5107

Title of paper: Empirical study of the topology and geometry of deep networks.

What is their primary result? The paper shows that state-of-the-art deep networks learn connected classification regions, and that the decision boundary in the vicinity of data points is flat along most directions. Moreover, they show that the decision boundary is biased towards negative curvatures; directions with significant curved decision boundaries are shared between data points; and they demonstrate that there is a relationship between the sensitivity of a classifier to perturbations and these shared directions.

Why is this important? It is important to understand the weaknesses of deep neural networks, specifically their instability under perturbation. This paper addresses the geometric properties of such deep networks to help shine a light on this problem.

What are their key ideas? Their key ideas are checking the connectivity of the network. In particular, the CaffeNet architecture on ImageNet classification. They find empirical evidence to conjecture that classification regions are connected in $\mathbb{R}^N$.

Next they consider the curvature of the LeNet and NiN architectures trained on the CIFAR-10 task, and show empirically that it is "almost" flat with a bias towards negative curvature.

What are the limitations, either in performance or applicability?

What might be an interesting next step based on this work?

What's the architecture? The architectures used are CaffeNet for the boundary regions and LeNet and NiN for the curvature analysis.

How did they train and evaluate it? These networks were trained on ImageNet, and CIFAR-10, respectively.

Did they implement something? The authors implemented two algorithms. Algorithm 1 finds a path between two datapoints, whereas Algorithm 2 detects image perturbation.

Grader's 598 ID: