ID de Contribution: **5** Type: **Non spécifié**

Input Similarity from the Neural Network Perspective

vendredi 5 février 2021 14:50 (40 minutes)

Given a trained neural network, we aim at understanding how similar it considers any two samples. For this, we express a proper definition of similarity from the neural network perspective (i.e. we quantify how undissociable two inputs A and B are), by taking a machine learning viewpoint: how much a parameter variation designed to change the output for A would impact the output for B as well?

We study the mathematical properties of this similarity measure, and show how to estimate sample density with it, in low complexity, enabling new types of statistical analysis for neural networks. We also propose to use it during training, to enforce that examples known to be similar should also be seen as similar by the network.

We then study the self-denoising phenomenon encountered in regression tasks when training neural networks on datasets with noisy labels. We exhibit a multimodal image registration task where almost perfect accuracy is reached, far beyond label noise variance. Such an impressive self-denoising phenomenon can be explained as a noise averaging effect over the labels of similar examples. We analyze data by retrieving samples perceived as similar by the network, and are able to quantify the denoising effect without requiring true labels.

Orateur: CHARPIAT, Guillaume (LRI)