

Born Again Neural Networks

Radek Bartyzal

Let's talk ML in Prague

December 18, 2017

Knowledge distillation

Transfer knowledge acquired by a learned teacher model to a new student model.

Teacher

- high-capacity model
- good performance

Student

- more compact model

By transferring knowledge, one hopes to benefit from the student's compactness while suffering only minimal degradation in performance.

Born Again Networks

Rather than compressing models, we train students that are parameterized identically to their parents. Surprisingly, these born again networks (BANs), tend to outperform their teacher models. Our experiments with born again dense networks demonstrate state-of-the-art performance on the CIFAR-100 dataset reaching a validation error of 15.5% with a single model and 14.9% with our best ensemble. Additionally, we investigate knowledge transfer to architectures that are different, but with capacity comparable to their teachers. In these experiments, we show that similar advantages can be achieved by transferring knowledge between dense networks and residual networks of similar capacity.



Tommaso Furlanello et al. "Born Again Neural Networks." Workshop on Meta-Learning (MetaLearn 2017) at NIPS. Accessible from: http://metalearning.ml/papers/metalearn17_furlanello.pdf