Paper Review (Lu Lu, Aug 16, 2019 @ Crunch Seminar)

Improving Simple Models with Confidence Profiles

Amit Dhurandhar*
IBM Research,
Yorktown Heights, NY
adhuran@us.ibm.com

Karthik Shanmugam
IBM Research,
Yorktown Heights, NY
karthikeyan.shanmugam2@ibm.com

Ronny Luss
IBM Research,
Yorktown Heights, NY
rluss@us.ibm.com

Peder Olsen
IBM Research,
Yorktown Heights, NY
pederao@us.ibm.com

NIPS 2018

Motivation

- A trained **deep** neural network that has a **high** test accuracy
- A **simpler** interpretable model or a very **shallow** network with a priori **low** test accuracy

Why?

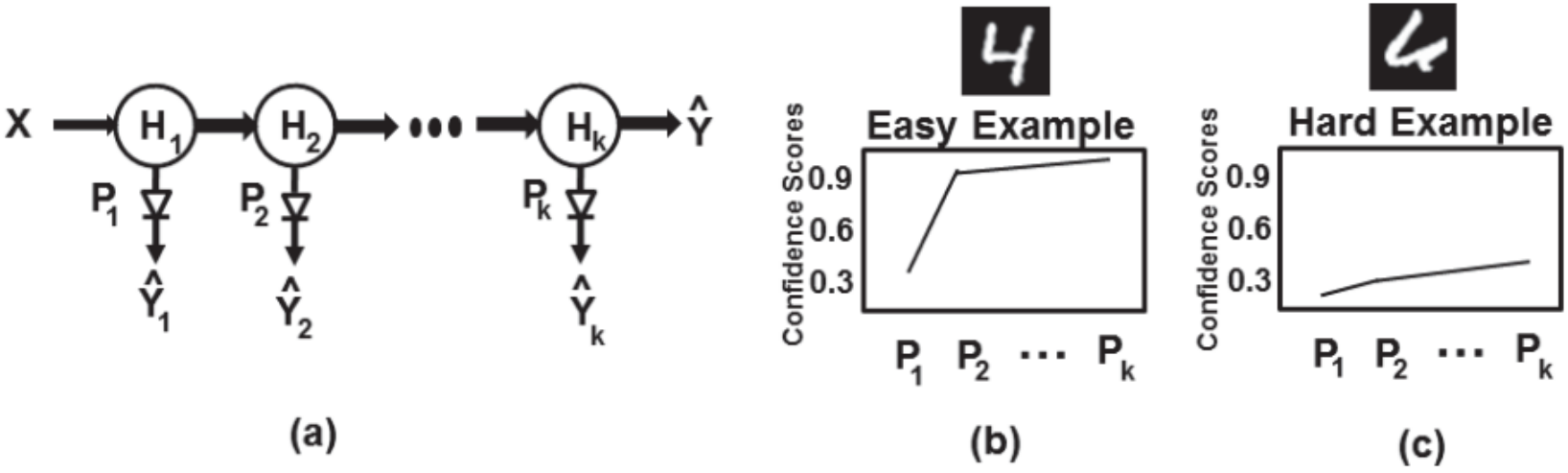
- Interpretability, e.g., medical decision
- Memory/power constrained, e.g., Internet-of-Things, mobile devices

Question:

- How to enhance the performance of simple models?

ProfWeight

- Add probes (logistic classifier, softmax($Wx + b$)) to the intermediate layers of a deep neural networks



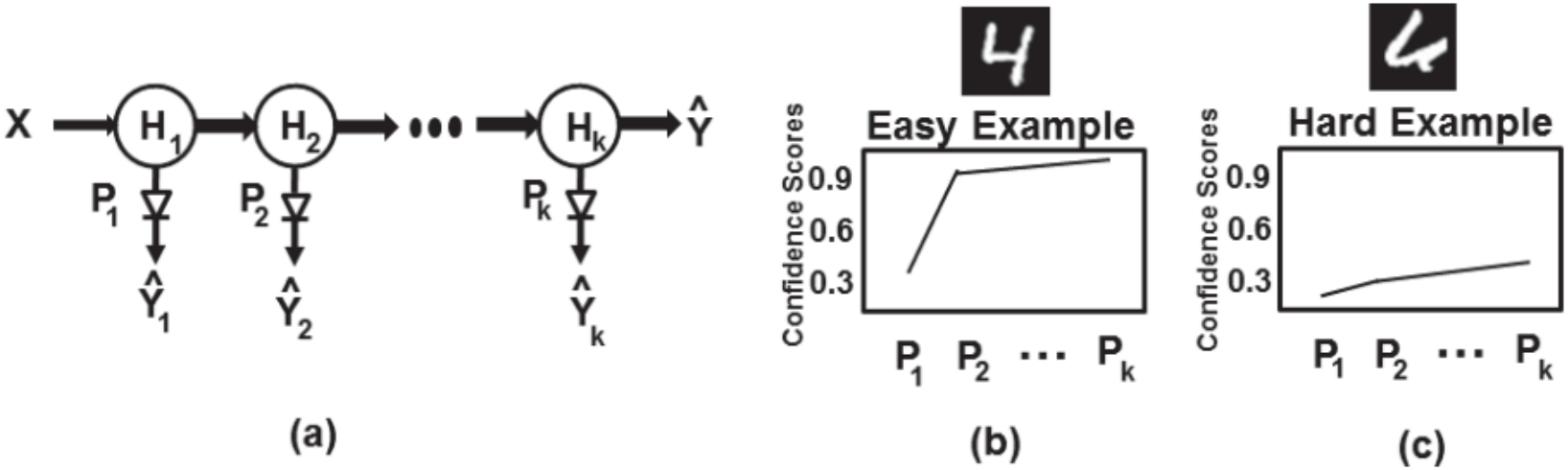
Algorithm

1. Train a deep network (no probes), and a simple model on a dataset.
2. Train probes.
3. For each data, learn its weight.
4. Train the simple model on the weighted dataset.

Weight computation I

Intuition: inform the simple model to ignore **hard** examples (**small** weight) and expend more effort on **easy** examples (**large** weight).

Confidence profile



- I: all probes that are more accurate than the simple model by a margin α
- AUC (area under the curve)

$$w_i = \frac{1}{|I|} \sum_{u \in I} c_{iu}$$

Weight computation II

$$S^* = \min_{\beta} \mathbb{E}[\lambda(S_\beta(x) - y)] \Rightarrow S^* = \min_{w \in \mathcal{C}} \min_{\beta} \mathbb{E}[\lambda(S_{w,\beta}(x) - y)]$$

$\mathcal{C}$ is a neural network: $c_{iu} \rightarrow w_i$

Algorithm

- Init weights $w = \mathbf{1}$.
- Loop
  - Update β, i.e., training the simple model S on the weighted dataset.
  - Update weights

$$w = \arg\min_{w \in \mathcal{C}} \mathbb{E}[\lambda(S_{w,\beta}(x) - y)] + \gamma \mathcal{R}(w)$$

$$\mathcal{R}(w) = (\frac{1}{m} \sum_i w_i - 1)^2$$

Experiments: CIFAR-10

- Complex model: ResNet with 15 blocks
- Simple models: ResNets with 3, 5, 7, and 9 blocks

	SM-3	SM-5	SM-7	SM-9
Standard	73.15(± 0.7)	75.78(±0.5)	78.76(±0.35)	79.9(±0.34)
ConfWeight	76.27 (±0.48)	78.54 (±0.36)	**81.46**(±0.50)	82.09 (±0.08)
Distillation	65.84(±0.60)	70.09 (±0.19)	73.4(±0.64)	77.30 (±0.16)
ProfWeight$^{\text{ReLU}}$	**77.52** (±0.01)	78.24(±0.01)	80.16(±0.01)	81.65 (±0.01)
ProfWeight$^{\text{AUC}}$	76.56 (±0.62)	**79.25**(±0.36)	**81.34**(±0.49)	**82.42** (±0.36)

- ConfWeight: w_i = the confidence score of the last probe

Experiments: Manufacturing dataset

Predict the quantity of metal etched on each wafer by 5104 inputs: acid concentrations, electrical readings . . .

- Complex model: FNN (5 hidden layers, 1024)
- Simple models: decision tree