Improving Simp

Amit Dhurandhar

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?

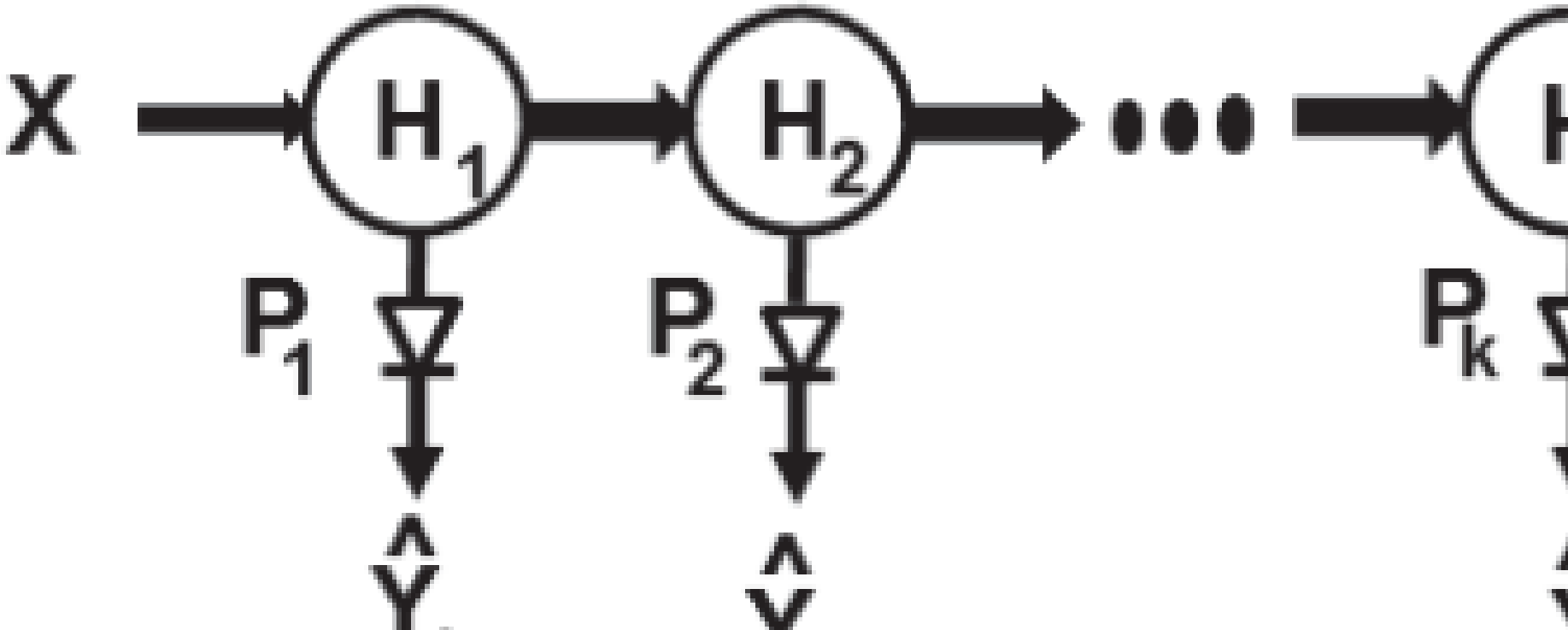
- Interpretability, e.g., medical decision
- Memory/power constrained, e.g., Internet-of-Things, mobile devices

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

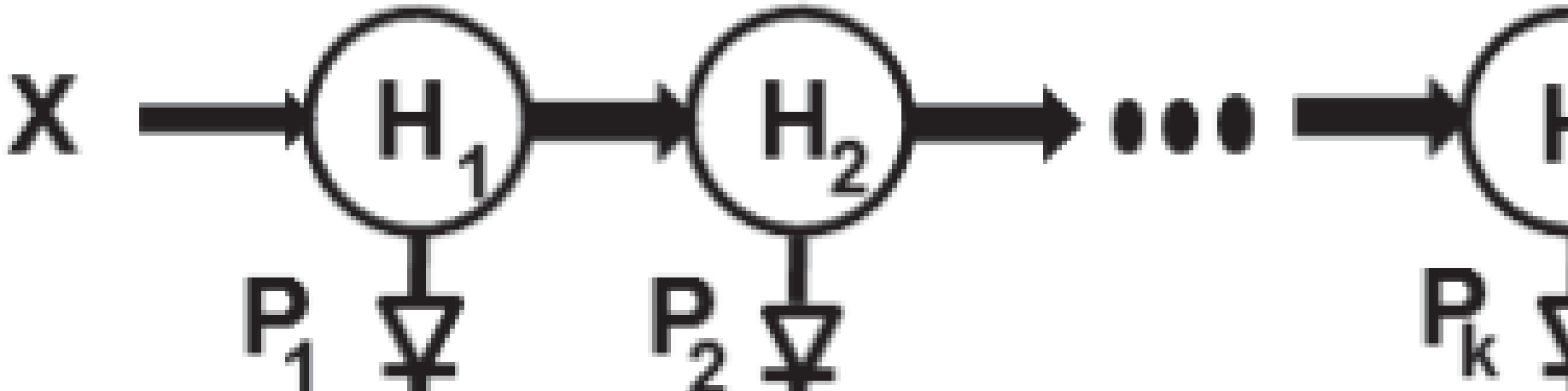


Weight computation I

Intuition

- Inform the simple model to ignore **hard** examples (**small** weight) and make it expend more effort on **easy** examples (**large** weight).

Confidence profile



Weight computation II

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

Algorithm

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

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

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

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

Experiments: CIFAR-10

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

	SM-3
Standard	73.15(± 0.7)
ConfWeight	76.27 (±0.48
Distillation	65.84(±0.60)
ProfWeight$^{\text{ReLU}}$	**77.52** (±0.01
ProfWeight$^{\text{AUC}}$	76.56 (±0.62

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