Dynamic Loss-Based

Parametric Curriculum

Inspired by: Saxena et al. Data Parameters: A New Family of Parameters for Learning a Differentiable Curriculum [NIPS'19]

Non-Parametric Curriculum

Inspired by: Castells et al. SuperLoss: A Generic Loss for Robust Curriculum Learning
[NIPS'20]

Parametric Curriculum

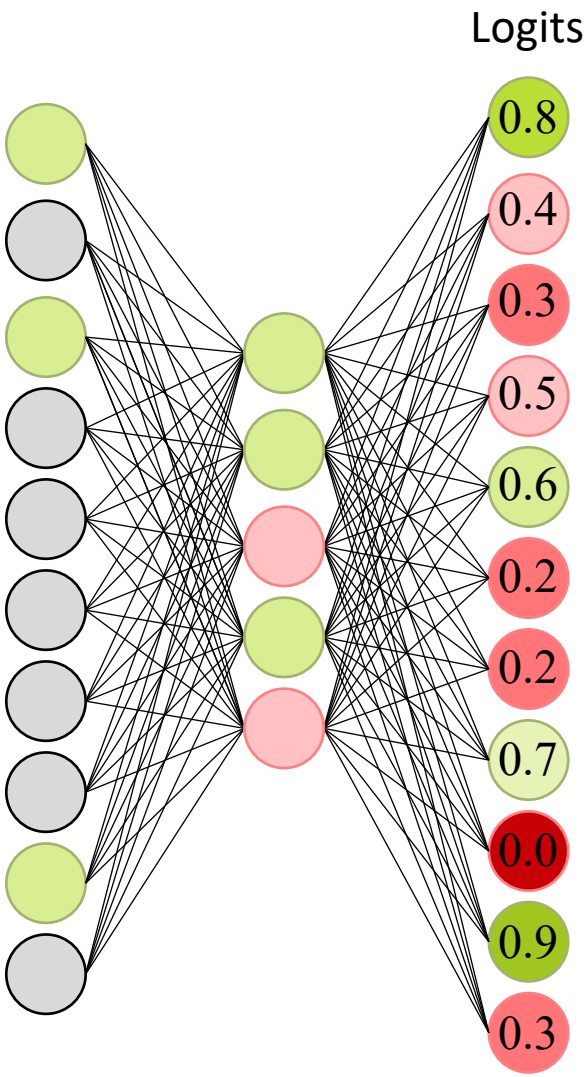
- A learnable parameter (ϕ) for each class (expert)
- $\frac{\text{Higher}}{\text{Lower}}$ ϕ value → $\frac{\text{More}}{\text{Less}}$ reliable output
- ϕ values change through the training process
- A class with high ϕ value score can be considered as an easy class

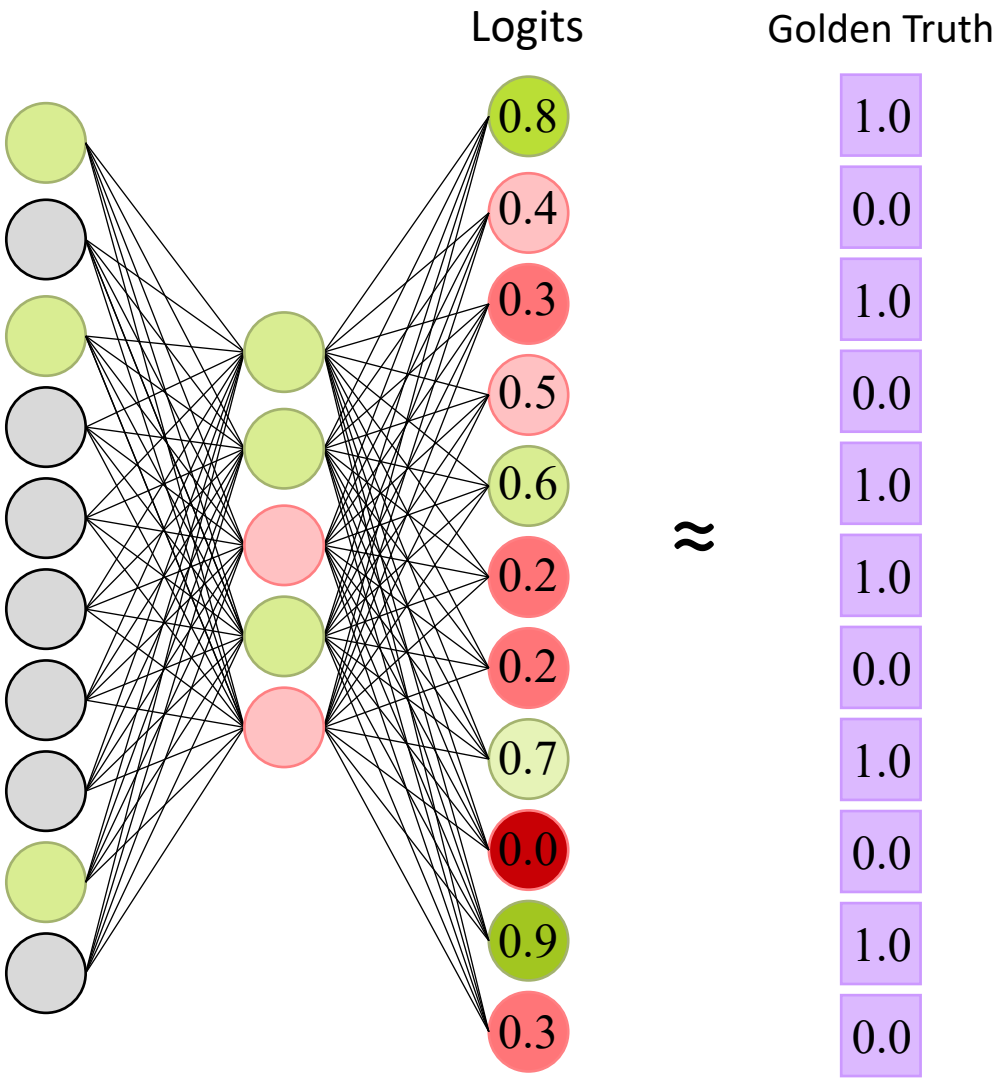
Binary Cross-Entropy

$$l(s,e) = \sum_{j \in e} -\log \sigma(z[j])$$

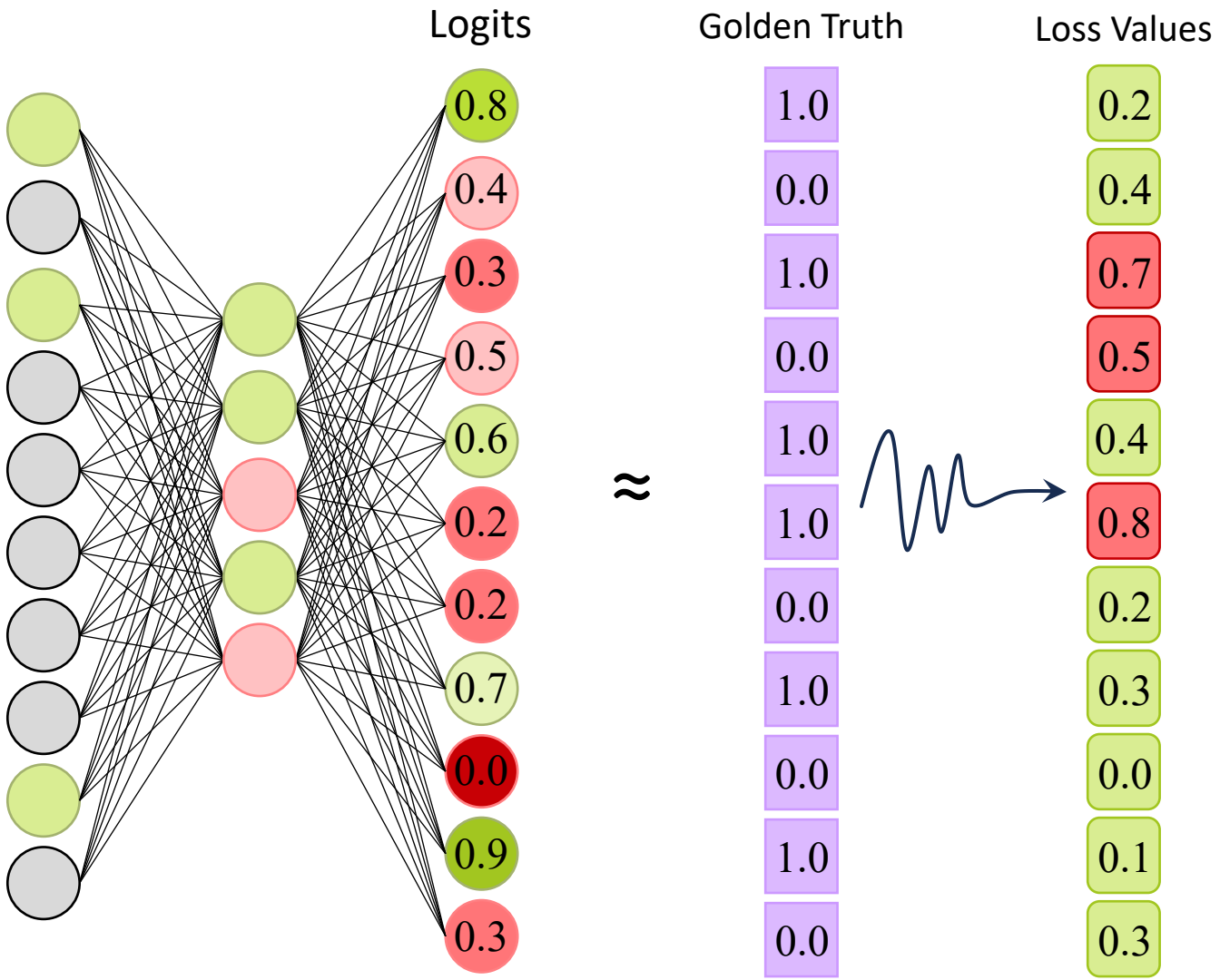
Proposed Binary Cross-Entropy

$$l\big((s,e),\phi\big) = \sum_{j \in e} -\log \sigma\left(\frac{z[j]}{\phi[j]}\right)$$

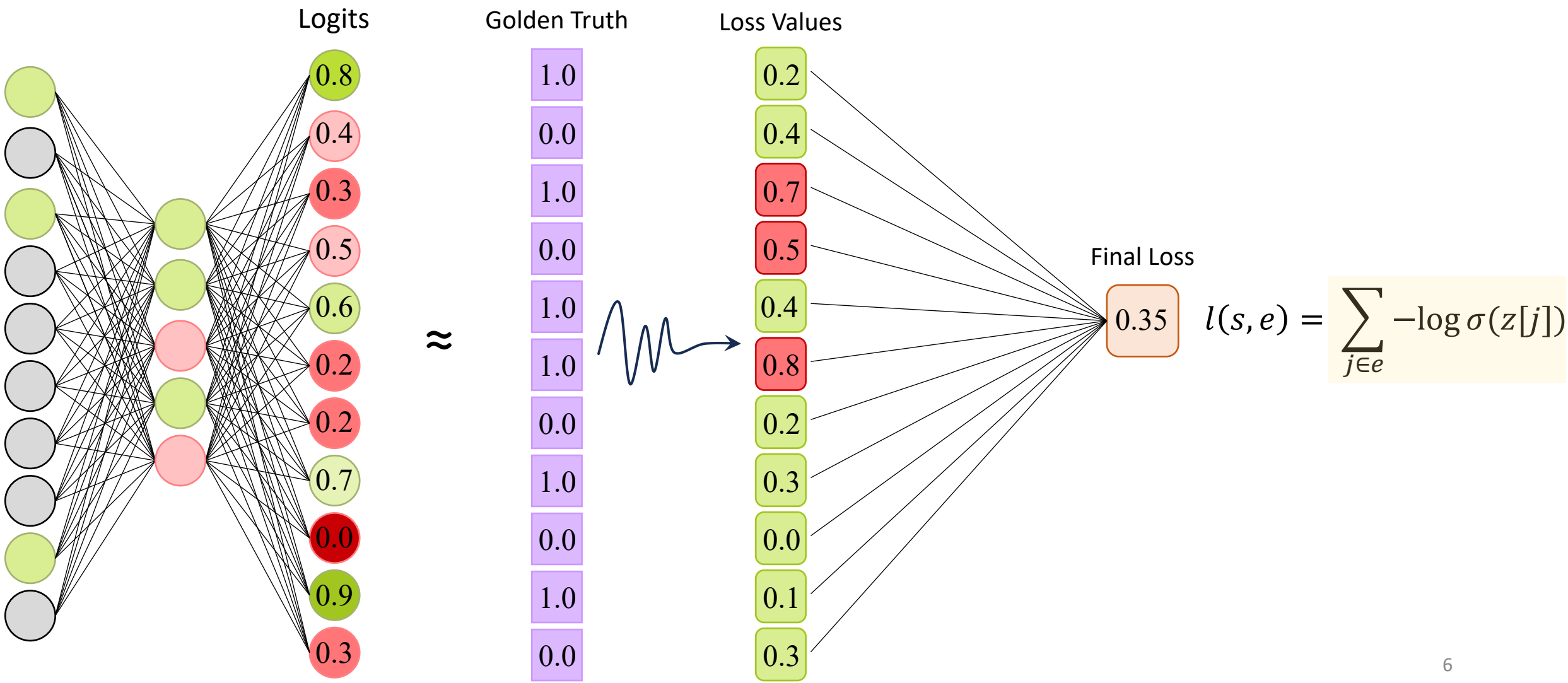

Logits
0.8
0.4
0.3
0.5
0.6
0.2
0.2
0.7
0.0
0.9
0.3

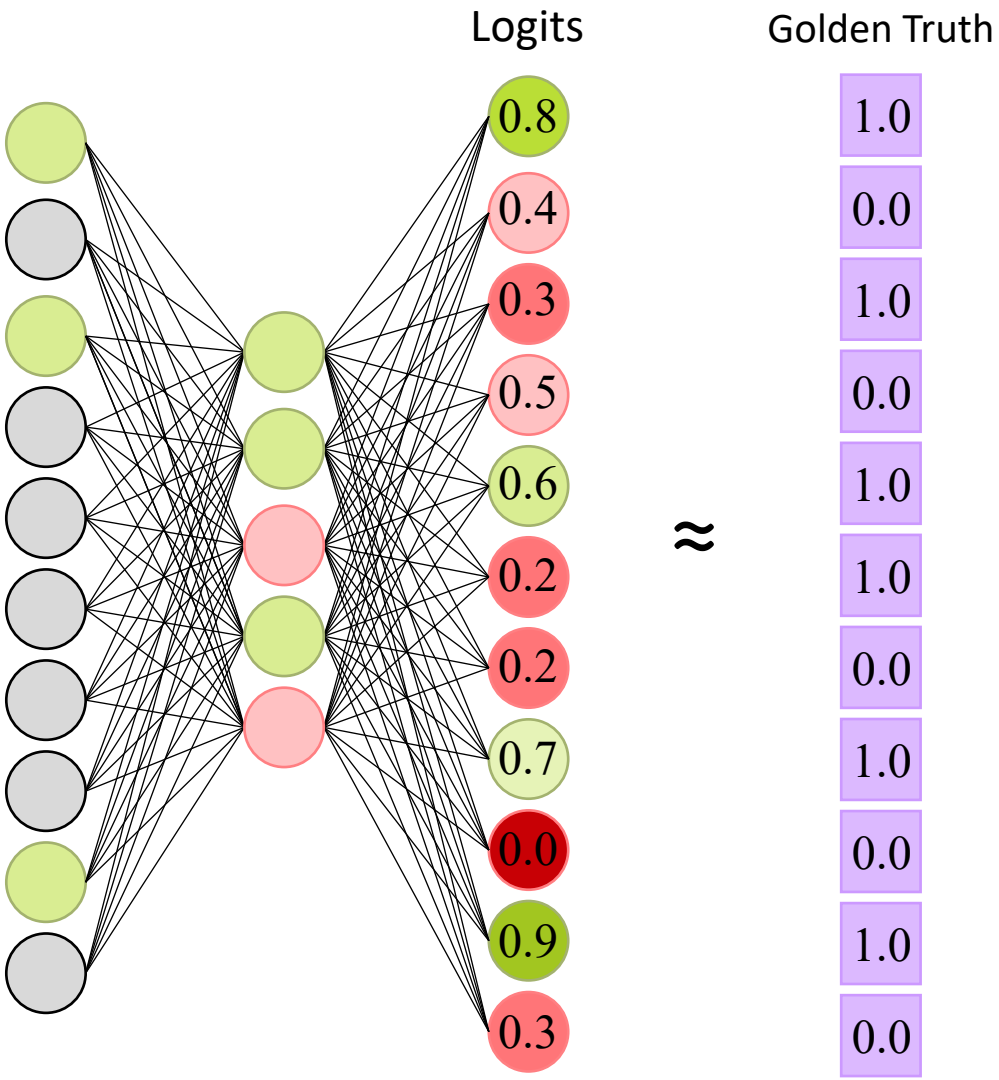


$$l(s,e) = \sum_{j \in e} -\log \sigma(z[j])$$

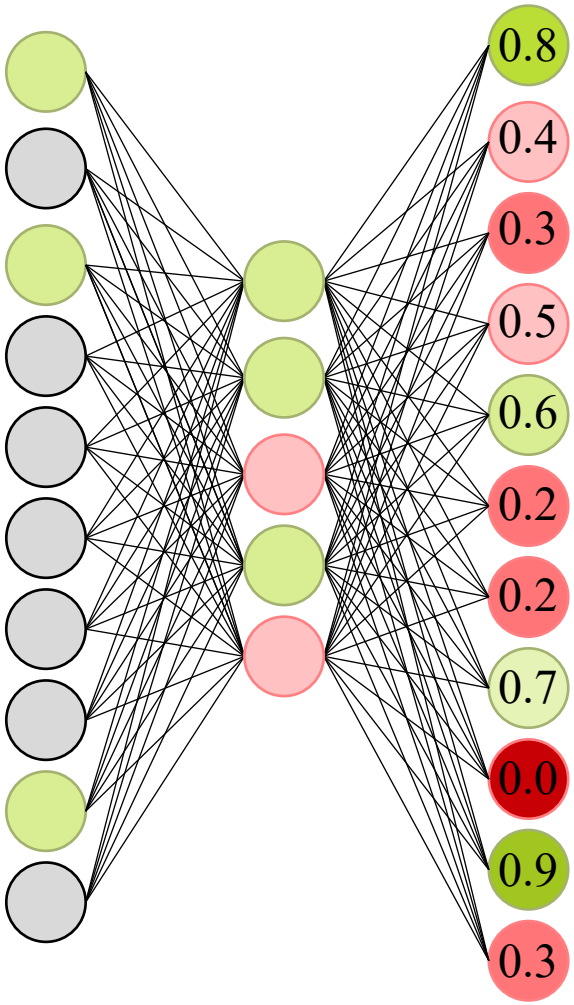


$$l(s,e) = \sum_{j \in e} -\log \sigma(z[j])$$


Logits
Golden Truth
Loss Values
0.8
0.4
0.3
0.5
0.6
0.2
0.2
0.7
0.0
0.9
0.3
≈
1.0
0.0
1.0
0.0
1.0
1.0
0.0
1.0
0.0
1.0
0.0
0.2
0.4
0.7
0.5
0.4
0.8
0.2
0.3
0.0
0.1
0.3
Final Loss
0.35
$l(s,e) = \sum_{j \in e} -\log \sigma(z[j])$



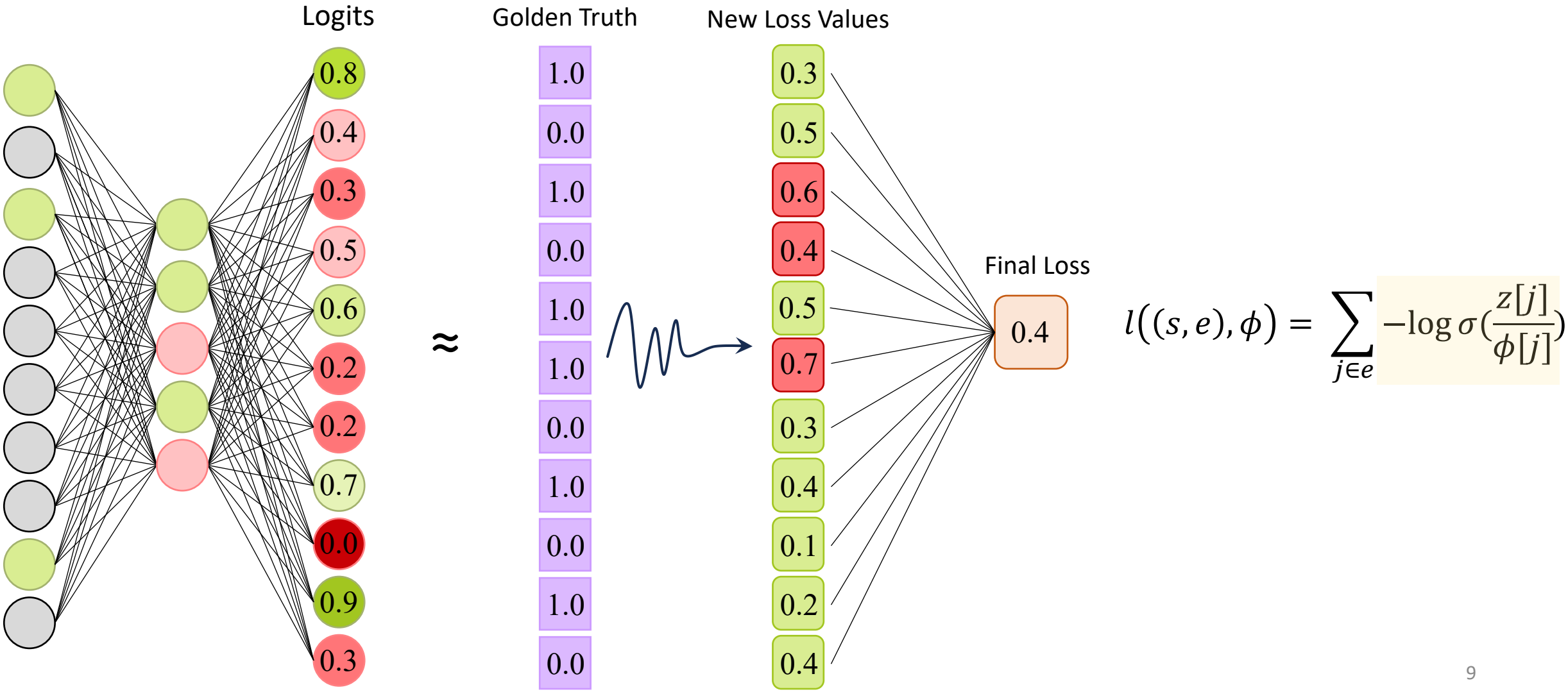
$$l\big((s,e),\phi\big) = \sum_{j \in e} -\log \sigma\left(\frac{z[j]}{\phi[j]}\right)$$



$\approx$

Logits	Golden Truth	New Loss Values		Old Loss Values
0.8	1.0	0.3	$high\ \phi$	0.2
0.4	0.0	0.5	$high\ \phi$	0.4
0.3	1.0	0.6	$low\ \phi$	0.7
0.5	0.0	0.4	$low\ \phi$	0.5
0.6	1.0	0.5	$high\ \phi$	0.4
0.2	1.0	0.7	$low\ \phi$	0.8
0.2	0.0	0.3	$high\ \phi$	0.2
0.7	1.0	0.4	$high\ \phi$	0.3
0.0	0.0	0.1	$high\ \phi$	0.0
0.9	1.0	0.2	$high\ \phi$	0.1
0.3	0.0	0.4	$high\ \phi$	0.3

$$l\big((s,e),\phi\big) = \sum_{j \in e} -\log \sigma\left(\frac{z[j]}{\phi[j]}\right)$$


Logits
Golden Truth
New Loss Values
Final Loss
0.8
0.4
0.3
0.5
0.6
0.2
0.2
0.7
0.0
0.9
0.3
≈
1.0
0.0
1.0
0.0
1.0
1.0
0.0
1.0
0.0
1.0
0.0
0.3
0.5
0.6
0.4
0.5
0.7
0.3
0.4
0.1
0.2
0.4
0.4
$l((s,e),\phi) = \sum_{j \in e} -\log \sigma(\frac{z[j]}{\phi[j]})$

Non-Parametric Curriculum

- Useing the converged values of ϕ at the limit

$$l\big((s,e),\phi[j]\big) = l((s,e),\phi^*[j])$$

- No learnable parameter for each class
- $\frac{\text{Higher}}{\text{Lower}}$ loss → $\frac{\text{Less}}{\text{More}}$ reliable output
- A class with $\frac{high}{low}$ loss can be considered as a $\frac{hard}{easy}$ one
- Amplify the effect of *easy* classes and reduce the effect of *difficult* classes on the final loss value