

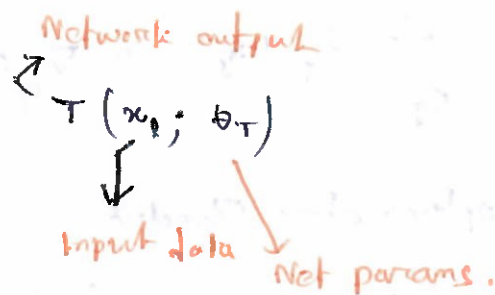
①

① meta pseudo labels: Meta pseudo-labels

Teacher, $T \rightarrow \theta_T$ Data labeled (x_L, y_L)

Student, $S \rightarrow \theta_S$

Notation Soft prediction



Pseudo label Optimization : (Review)

$$\theta_S^{PL} = \arg \min_{\theta_S} \mathbb{E}_{x_u} \left[\text{CE} \left(T(x_u; \theta_T), S(x_u; \theta_S) \right) \right]$$

$$:= \mathcal{L}_u(\theta_T, \theta_S)$$

$$\mathbb{E}_{x_L, y_L} \left[\text{CE} \left(y_L, S(x_L; \theta_S^{PL}) \right) \right] := \mathcal{L}_L(\theta_S^{PL}) \quad // \text{should be low}$$

↓
is a function of (θ_T)

further,

$$\min_{\theta_T} \mathcal{L}_L(\theta_S^{PL}(\theta_T))$$

$$\text{where } \theta_S^{PL}(\theta_T) = \arg \min_{\theta_S} \mathcal{L}_u(\theta_T, \theta_S)$$

pseudo label adjustment 'is possible in

new optimization problem without everything!!

(11)

Practical Approximation:

$$\theta_s^{PT}(\theta_T) \approx \theta_s - \eta_s \nabla_{\theta_s} L_u(\theta_T, \theta_s)$$

$$\min_{\theta_T} L_L(\theta_s - \eta_s \nabla_{\theta_s} L_u(\theta_T, \theta_s))$$

SGD optimization Objective:

$$\theta_s' = \theta_s - \eta_s \nabla_{\theta_s} L_u(\theta_T, \theta_s) \quad // 1st$$

$$\theta_T' = \theta_T - \eta_T \nabla_{\theta_T} L_L(\underbrace{\theta_s - \nabla_{\theta_s} L_u(\theta_T, \theta_s)}_{\text{labeled data}}) \quad // 2nd.$$

unlabeled