3 FedNTD：联合非真实蒸馏****

> FedNTD 的核心思想是只为非真实类保留全局视图。
>
> 进行局部蒸馏，通过交叉熵损失函数
>
> $$\mathcal{L} = \mathcal{L}_{\mathrm{CE}}(q^l,\ \mathbb{1}_y) + \beta \cdot \mathcal{L}_{\mathrm{NTD}}(\tilde{q}_\tau^l,\ \tilde{q}_\tau^g).$$
>
> **Algorithm 1** Federated Not-True Distillation (FedNTD)
>
> **Input:** total rounds T, local epochs E, dataset $\mathcal{D}$, sampled clients sets $K^{(t)} \subset K$ in round t, learning rate γ
>
> **Initialize** $w^(0)$ for global server weight
>
> **for** each communication round $t = 1, \cdots, T$ **do**
>
> Server samples clients $K^{(t)}$ and broadcasts $\tilde{w}^{(t)} \leftarrow w^{(t)}$
>
> **for** each client $k \in K^{(t)}$ **in parallel do**
>
> **for** Local Steps $e = 1 \cdots E$ **do**
>
> **for** Batches $j = 1 \cdots B$ **do**
>
> $\tilde{w}_k^{(t)} \leftarrow \tilde{w}_k^{(t)} - \gamma \nabla_w \mathcal{L}(\tilde{w}_k^{(t)}; [\mathcal{D}^k]_j)$ Using [Equation 10]
>
> **end for**
>
> **end for**
>
> **end for**
>
> Upload $\tilde{w}_k^t$ to server
>
> **Server Aggregation :** $w^{(t+1)} \leftarrow \frac{1}{|K^{(t)}|} \sum_{k \in K^{(t)}} \tilde{w}_k^{(t)}$
>
> **end for**
>
> **Server output :** w_T
>
> 与普通的联邦学习的唯一不同就是
>
> 在每个batches更新全局模型时，所采用的目标损失函数不同。