Algorithm 2: Federated learning with Proximal regularization eXcept local Normalization (`FedPxN`)

Notation: number of clients K, number of communication rounds T, number of local epochs E, Data $D := (D_1, D_2, ..., D_K)$, learning rate η, normalization layers *norm*

Server executes:

initialize model parameters w_0

for $t = 0, ..., T-1$ **do**

  for *each client* $k \in K$ **do**

    $w_{t,k \setminus norm} \leftarrow w_{t \setminus norm}$

    $w_{t,k} \leftarrow \text{LocalTraining}(k,\ w_{t,k},\ D_k)$

  end

  $w_{t+1 \setminus norm} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n}\ w_{t,k \setminus norm}$

end

LocalTraining(k, $w_{t,k}$, D_k):

for $e = 0, ..., E-1$ **do**

  for *batch* $b \leftarrow$ *(x, y) of* D_k **do**

    $R = \|w_{t,k \setminus norm} - w_{t \setminus norm}\|^2$

    $w_{t,k} \approx \arg\min_{w} F_k(w; b) + \frac{\mu}{2} R$

  end

end

return $w_{t,k}$