$$\alpha_k = \left\|\theta_t - \theta_t^{(k)}\right\|^{-1} / (\sum_{k=1}^{K} \left\|\theta_t - \theta_t^{(k)}\right\|^{-1}).$$

Considering the time effect during federated communication, Chen et al. [2020] proposed temporally weighted aggregation of the local models on the server as:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} (\frac{e}{2})^{-(t-t^{(k)})} \theta_t^{(k)},$$

where e is the natural logarithm, t is the current update round and $t^{(k)}$ is the update round of the newest $\theta^{(k)}$.

Attentive Aggregation

The federated averaging algorithm takes the instance ratio of the client as the weight to calculate the averaged neural parameters during model fusion [McMahan et al., 2017]. In attentive aggregation, the instance ratio is replaced by adaptive weights as Eq. 5:

$$\theta_{t+1} \leftarrow \theta_t - \epsilon \sum_{k=1}^{m} \alpha_k \nabla \mathcal{L}(\theta_t^{(k)}),$$