

第 2 章：联邦学习基础算法

思考题 2.1

(a) 联邦平均以各客户端数据量占比为权重进行加权聚合：

$$\theta_{\text{global}} = \frac{|\mathcal{D}_1|}{|\mathcal{D}|} \theta_1 + \frac{|\mathcal{D}_2|}{|\mathcal{D}|} \theta_2 = \frac{200}{500} \times 1.8 + \frac{300}{500} \times 2.2 = 0.72 + 1.32 = 2.04.$$

结果略偏向 Client B，这是因为 Client B 的数据量更多，权重更大。

(b) 两个客户端均从 $\theta_{\text{global}} = 2.0$ 出发，执行 $E = 2$ 步本地梯度下降，更新规则为 $\theta \leftarrow \theta - \eta g$ 。

Client A ($g_1 = -0.4$) 两步后：

$$\theta_1^{\text{new}} = 2.0 - 2 \times 0.1 \times (-0.4) = 2.0 + 0.08 = 2.08.$$

Client B ($g_2 = 0.2$) 两步后：

$$\theta_2^{\text{new}} = 2.0 - 2 \times 0.1 \times 0.2 = 2.0 - 0.04 = 1.96.$$

聚合后的全局参数：

$$\theta_{\text{global}}^{\text{new}} = \frac{200}{500} \times 2.08 + \frac{300}{500} \times 1.96 = 0.832 + 1.176 = 2.008.$$

(c) 设当前全局参数为 $\theta^{(t)}$ ，客户端 k 在本地执行 E 步梯度下降后得到：

$$\theta_k^{\text{new}} = \theta^{(t)} - E\eta g_k.$$

对其按数据量加权聚合：

$$\begin{aligned} \theta^{(t+1)} &= \sum_k \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \theta_k^{\text{new}} = \sum_k \frac{|\mathcal{D}_k|}{|\mathcal{D}|} (\theta^{(t)} - E\eta g_k) \\ &= \theta^{(t)} - E\eta \underbrace{\sum_k \frac{|\mathcal{D}_k|}{|\mathcal{D}|} g_k}_{\bar{g}}. \end{aligned}$$

这与以有效步长 $E\eta$ 对全局加权梯度 \bar{g} 做一次梯度下降完全等价，即

$$\theta^{(t+1)} = \theta^{(t)} - E\eta \bar{g}, \quad \bar{g} = \sum_k \frac{|\mathcal{D}_k|}{|\mathcal{D}|} g_k.$$

由全局损失 $\mathcal{L}(\theta) = \frac{|\mathcal{D}_1|}{|\mathcal{D}|} \mathcal{L}_1(\theta) + \frac{|\mathcal{D}_2|}{|\mathcal{D}|} \mathcal{L}_2(\theta)$ 可知 $\nabla \mathcal{L} = \bar{g}$ ，故 FedAvg 的聚合方向与最小化全局损失函数的梯度方向完全一致。

对于数据量差异的影响，数据量较大的客户端在 \bar{g} 中占据更高权重，其本地数据分布对全局模型的影响更显著。当各客户端数据分布存在较大差异 (Non-IID) 时，数据量少的客户端容易被“压制”，导致全局模型对其欠拟合。实践中常通过引入正则化约束（如 FedProx）或自适应权重来缓解这一问题。

思考题 2.2

(a) 设备 A 利用同态加密的线性运算性质, 在密文上直接计算加密预测值。对第 i 个样本:

$$\llbracket \hat{y}_i \rrbracket = w_A x_i^A + w_B \llbracket s_i^B \rrbracket + b.$$

代入 $w_A = 0.8$, $w_B = 0.6$, $b = 10$ 及各样本数据:

$$\begin{aligned}\llbracket \hat{y}_1 \rrbracket &= 0.8 \times 25 + 0.6 \times \llbracket 5 \rrbracket + 10 = \llbracket 33 \rrbracket, \\ \llbracket \hat{y}_2 \rrbracket &= 0.8 \times 30 + 0.6 \times \llbracket 7 \rrbracket + 10 = \llbracket 38.2 \rrbracket, \\ \llbracket \hat{y}_3 \rrbracket &= 0.8 \times 35 + 0.6 \times \llbracket 9 \rrbracket + 10 = \llbracket 43.4 \rrbracket.\end{aligned}$$

上述运算均在密文空间中完成, 设备 A 始终无法获知 s_i 的明文值。

(b) 损失函数为 $\mathcal{L} = \frac{1}{6} \sum_{i=1}^3 (y_i - \hat{y}_i)^2$, 其中 $\hat{y}_i = w_A x_i + w_B s_i + b$ 。对 w_A 和 b 求偏导, 得到梯度公式:

$$\frac{\partial \mathcal{L}}{\partial w_A} = -\frac{1}{3} \sum_{i=1}^3 (y_i - \hat{y}_i) x_i, \quad \frac{\partial \mathcal{L}}{\partial b} = -\frac{1}{3} \sum_{i=1}^3 (y_i - \hat{y}_i).$$

将已知的残差代入 ($y_1 - \hat{y}_1 = 50 - 50.5 = -0.5$, $y_2 - \hat{y}_2 = 70 - 70.3 = -0.3$, $y_3 - \hat{y}_3 = 90 - 90.1 = -0.1$):

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_A} &= -\frac{1}{3} [(-0.5) \times 25 + (-0.3) \times 30 + (-0.1) \times 35] \approx 8.33, \\ \frac{\partial \mathcal{L}}{\partial b} &= -\frac{1}{3} [(-0.5) + (-0.3) + (-0.1)] = -\frac{-0.9}{3} = 0.30.\end{aligned}$$

(c) 设备 A 按梯度下降规则更新参数 ($\eta = 0.01$):

$$\begin{aligned}w_A^{\text{new}} &= w_A - \eta \frac{\partial \mathcal{L}}{\partial w_A} = 0.8 - 0.01 \times 8.33 \approx 0.717, \\ b^{\text{new}} &= b - \eta \frac{\partial \mathcal{L}}{\partial b} = 10 - 0.01 \times 0.30 = 9.997.\end{aligned}$$

整个过程中, 设备 B 的参数 w_B 由设备 B 自行利用解密后的残差梯度独立更新, 两方无需交换明文数据。

思考题 2.3

对设备 n 的参数更新量取条件期望 (给定 w_t^n):

$$\mathbb{E}[w_{t+1}^n - w_t^n \mid w_t^n] = -\eta \mathbb{E}[\nabla l_A(w_t^n; D_n^A) + \nabla l_B(w_t^n; D_n^B)] + \frac{\gamma}{N-1} \sum_{m \neq n} \mathbb{E}[\nabla l_A(w_t^m; D_m^A) + \nabla l_B(w_t^m; D_m^B)].$$

记 $\Delta^n \triangleq \mathbb{E}[w_{t+1}^n - w_t^n \mid w_t^n]$, 类似地记 Δ^{n_1} 和 Δ^{n_2} 。

要证 $\Delta^{n_1} \perp \Delta^{n_2}$ ，只需证明 $\mathbb{E}[\Delta^{n_1} \cdot \Delta^{n_2}] = 0$ 。

注意到 Δ^{n_1} 由设备 n_1 自身的梯度及其他设备（含 n_2 ）的梯度期望组成， Δ^{n_2} 同理。对乘积展开后，交叉项均形如

$$\mathbb{E}[\nabla l_A(w; D_{n_1}^A) \cdot \nabla l_A(w; D_{n_2}^A)] \quad \text{或} \quad \mathbb{E}[\nabla l_B(w; D_{n_1}^B) \cdot \nabla l_B(w; D_{n_2}^B)].$$

由题目假设，对任意 $n_1 \neq n_2$ 及任意 w ，这两类期望均为零。因此所有涉及 n_1 与 n_2 数据的交叉项消失，从而：

$$\mathbb{E}[\Delta^{n_1} \cdot \Delta^{n_2}] = 0,$$

即两设备的条件期望参数更新量相互正交（独立），命题得证。 \square

思考题 2.4

(a) 各节点不应直接共享验证损失的原始数值，而应共享各 λ 下损失的**排名**（即节点内部按损失从小到大对候选 λ 的排序结果）。排名信息不会暴露节点的实际数据分布或损失量级，同时足以用于全局决策。

若需更强的隐私保证，也可对损失值加入拉普拉斯噪声（满足差分隐私）后再共享，或使用安全多方计算对损失求和而不暴露各节点的具体数值。

(b) 收集各节点共享的信息后，计算每个 λ 候选值的整体平均验证损失：

$$\bar{\mathcal{L}}(\lambda) = \frac{1}{K} \sum_{k \in \{A, B, C\}} \mathcal{L}_k(\lambda),$$

选取使 $\bar{\mathcal{L}}(\lambda)$ 最小的 λ^* 作为全局最优正则化参数：

$$\lambda^* = \arg \min_{\lambda \in \{0.01, 0.1, 1\}} \bar{\mathcal{L}}(\lambda).$$

(c) 各 λ 的整体平均验证损失：

$$\begin{aligned} \bar{\mathcal{L}}(0.01) &= \frac{0.25 + 0.30 + 0.28}{3} = \frac{0.83}{3} \approx 0.277, \\ \bar{\mathcal{L}}(0.1) &= \frac{0.20 + 0.23 + 0.21}{3} = \frac{0.64}{3} \approx \mathbf{0.213}, \\ \bar{\mathcal{L}}(1) &= \frac{0.35 + 0.33 + 0.37}{3} = \frac{1.05}{3} = 0.350. \end{aligned}$$

$\lambda = 0.1$ 对应的平均验证损失最低，因此**最优正则化参数为 $\lambda^* = 0.1$** 。这一结论在三个节点中也完全一致——每个节点单独来看， $\lambda = 0.1$ 均为其本地最优选择，说明该参数具有良好的跨节点泛化性。