

FL教案·丸一口

所有不明白的名词都需要额外地查阅 博客、GPT 去理解；

每条大任务结束之后，都应该有师兄/姐去开会考察（主要是问一些列出来的“思考问题”之外的东西，因为“思考问题”是很容易用GPT生成的，这种开会考察的考核者是要比较熟悉这些内容的），顺便指点迷津。

1.联邦学习基础知识（3天）

1.1 视频教程

【【前置知识】FedAvg：联邦学习创世纪算法】 [🔗 https://www.bilibili.com/video/BV1Ke4y1r7ph/](https://www.bilibili.com/video/BV1Ke4y1r7ph/)

1.2 论文原文

Communication-Efficient Learning of Deep Networks from Decentralized Data

1.3 思考问题

1. 联邦学习与传统机器学习的区别
2. 各客户端聚合权重如何分配
3. local epoch的作用
4. 数据异质指的是什么，实验仿真的时候是如何为各个客户端设置异质数据的
5. 除了FedAvg中使用的病理性(Pathological)异构，还有哪种主流的数据异质划分方式？（需
| 额外查阅资料）

2. PFLlib仓：数据划分（4天）

[🔗 https://github.com/TsingZ0/PFLlib](https://github.com/TsingZ0/PFLlib)

2.1 视频教程

【【代码带读】PFL「壹」数据集处理】 [🔗 https://www.bilibili.com/video/BV1bu411K7bV/](https://www.bilibili.com/video/BV1bu411K7bV/)

2.2 思考问题

1. CV(计算机视觉)的数据集 MNIST、FashionMNIST、Cifar10、Cifar100，是长什么样的
2. NLP(自然语言处理)的数据集 AG news 是长什么样的
3. 如何通过修改参数，调整IID、病理性异构、Dirichlet 数据分布的生成
4. `X, y, statistic = separate_data()` 函数做了什么事
5. `train_data, test_data = split_data(X, y)` 函数做了什么事

3. PFLlib仓: Client 和 Server (4天)

3.1 视频教程

【【代码带读】PFL「贰」Client View】 <https://www.bilibili.com/video/BV1zN4y1y76b>

【【代码带读】PFL「叁」Server View】 <https://www.bilibili.com/video/BV1ew411r7Hn>

3.2 思考问题

1. `model.train()`和`model.eval()`是什么作用
2. 真正的模型更新在代码的哪个部分
3. 如何在`loss.backward()`之后马上拿到loss的数值（需要查阅额外资料，请把loss值打印在控制台）
4. 当需要采样客户端的时候，客户端权重是如何计算的；为什么要这样计算
5. 目前使用的是SGD optimizer：（需查阅额外资料）
 - a. 什么是optimizer
 - b. 如何更换SGD优化器为Adam优化器
6. 目前用的神经网络是什么结构
 - a. 这样一个网络，各层作用是什么
 - b. 为什么要用激活函数

4. 期末考（3天）

引子问题： PFLlib计算最终Accuracy的时候是如何计算的？

引子答案： 各客户端计算本地的Accuracy，再在中心方加权，得到最终Accuracy

终极任务： PFLlib的准确率计算方式是面向个性化联邦（Personalized Federated Learning, PFL）的，传统联邦学习的准确率计算方式是在中心方拥有测试集，并在中心方使用测试集对聚合模型进行准确率测试。

请向PFLlib加入中心方测试集生成的代码，一个可行的思路如下：

1. 将所有客户端的测试集合并，生成Server方的测试集
2. 在获得聚合模型后，使用生成的Server测试集对聚合模型进行准确率测试