隐私保护大作业：FedAvg 算法

一、 实验目的

掌握联邦学习基本技能，理解 FedAvg 算法，并且探究数据非独立同分布、模型稀疏更新情况下的算法收敛情况。

二、 实验步骤

1. 在 MNIST 上实现基础的 FedAvg 算法

提供的 Demo 里已经有 FedAvg 的基本实现，且在 Examples/MNIST 文件夹下有 DNN 和 CNN 两种网络的定义。

2. 考察不同的客户端比例

考察选取的客户端数目不同的联邦学习收敛情况。客户端总数固定 100 个，此时考察每次 FedAvg 更新时选取的客户端数量不同的情况，如：5 个，10 个，30 个，100 个。绘制不同客户端数目下的测试集 accuracy 曲线。

3. 考察不同数据分布

考察数据分布不同的情况下的联邦学习收敛情况。Demo 里面已经提供了两种客户端数据生成方式。1. IID，表示随机分割，各个客户端数据分布一致；2. 各个客户端仅包含某几个类别，而非所有的 10 个类别。每个类别包含的样本数量相等。

绘制 IID 情况以及客户端分别包含 1、2、3、4 类都样本情况下的测试集 accuracy 曲线。

4. 考察稀疏更新的情况。

稀疏更新，指的是每轮客户端上传**稀疏的模型更新量**，而非完整的模型或模型更新量。考察采用 top-k 稀疏（模型参数更新量按照绝对值排序，只选取前 k 的比例，其他变成 0）的情况，绘制 k = 1/1000, 1/500, 1/100, 1/50, 1/10 的 accuracy 曲线图

三、 提交内容

实验报告：包含上述实验内容的 pdf 文件。

实验代码：删除数据文件以及其他临时文件（如.idea、__pycache__）的实验代码，仅包含 python 文件。

参考文献：

差分隐私学习资料：The Algorithmic Foundations of Differential Privacy

稀疏更新论文：Sparse Communication for Distributed Gradient Descent