纵向联邦学习(VFL)的HE-DP隐私保护框架

作品源码github地址：https://github.com/XtangEver/HE-DP

作品docker镜像源码地址：docker pull tangxiandedog/fate:last

理论分析

在多方场景下的纵向联邦学习中，目前使用的隐私安全计算技术有同态加密(HE)、差分隐私(DP)等。在对逻辑回归、线性回归、泊松回归等广义线性模型的分析中，我们团队发现基于同态加密(HE)的方案中有存在**隐私泄露**风险的地方，发现现有的基于差分隐私(DP)的方案中由于需要对每条数据、损失函数添加噪音，导致全局敏感度较大进而影响到**噪音规模太大**进而影响精确度的问题。

所以，我们提出了基于HE-DP的隐私保护框架。**核心思想**是使用同态加密技术保证计算过程的安全性，使用差分隐私技术保证计算结果的安全性。详细来说：同态加密(HE)解密后的梯度由于含有对方的数据信息，存在隐私泄露的风险，故我们在密文梯度上使用差分隐私(DP)技术扰动数据，保护用户数据信息。

由于广义线性模型(GLM)在纵向联邦学习交互中具有类似性，并且密文梯度具有线性可分性。所以我们后续就仅以**纵向逻辑回归为例**(三个广义线性模型的算法实现放在对应的github地址上)。我们依次分析使用同态加密技术(HE)、差分隐私技术的流程细节，最后提出基于HE-DP的方案。

基于同态加密(HE)技术

在这里多方我们采用和fate框架中类似的表述，主要如下：

> guest:数据应用方，在纵向FL中指的是含有数据标签的一方。而且建模流程一般由guest方发起
>
> host：数据的提供方
>
> arbiter:主要来辅助多方联合建模的。纵向FL中用来分发公私钥、加解密服务等等。它本身既不提供数据，也不使用数据

首先要明确最终的目标：最终我要得到两个子模型，对于guest方(为了公式描述方便，以下也叫做A方)而言，我要得到子模型w_A；对于host方而言，我要得到子模型(为了公式描述方便，以下也叫做B方)ω_B

这也意味着当我使用模型进行预测的时候，我需要联合两个子模型ω_A和 ω_B才可以进行预测.

损失函数

$$F(\omega) = \frac{1}{n}\sum_{i=1}^{n} log(1+e^{-y_i\omega^T x_i})$$

损失函数的梯度公式是

$$\nabla F(\omega) = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{1+e^{-y_i\omega^T x_i}} - 1)y_i x_i$$

由于同态加密不支持log运算，所以需要将上述的损失函数形式近似转化为**多项式**才方可.

所以损失函数的梯度公式可以**近似**的写成

$$\nabla F(\omega) \approx \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x_i - 0.5y_i)x_i$$

为了方便表示,我们令

$$0.25\omega^T x_i - 0.5y_i = d_i$$

于是,损失函数梯度公式可以表示成

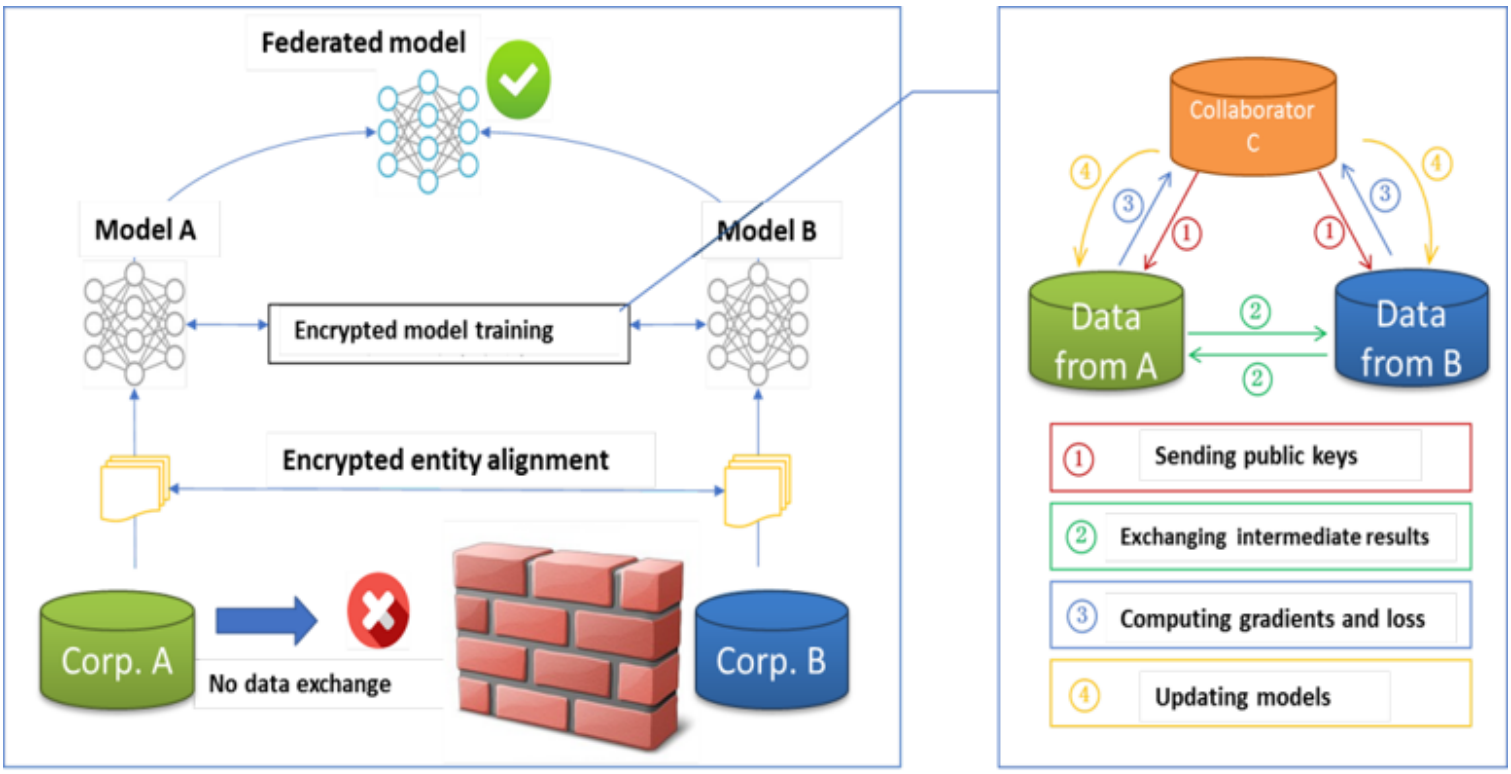
$$\nabla F(\omega) \approx \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x_i - 0.5y_i)x_i = \frac{1}{n}\sum_{i=1}^{n} d_i x_i$$

在纵向联邦学习（使用同态加密技术）的场景下，guest方拥有数据X_A，以及数据标签Y。host方(为了公式描述方便，以下也叫做B方)拥有数据X_B

所以完整的$\omega X_A = \omega_A X_A + \omega_B X_B$

这里有一个符号表示：$\omega_A X_A$当使用同态加密形式，则表示为：$[[\omega_A X_A]]$

算法执行流程(粗略描述)



a b

1. arbiter方分发公钥给A方和B方
2. B方计算$\omega_B X_B$，然后加密为$[[\omega_B X_B]]$。并且发给A方
3. A方收到$[[\omega_B X_B]]$后，同时自身也计算$[[\omega_A X_A]]$。然后获得完整的：$[[\omega x]] = [[\omega_A X_A]] + [[\omega_B X_B]]$
4. 这样，A方便获得了完整的损失函数。然后，将损失函数发送给B方

经过上述的4个步骤，双方便分别确定了下式（密文状态下）：

对于A方而言：

$$\nabla F(\omega)_A = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_A = \frac{1}{n}\sum_{i=1}^{n} d_i x_A$$

对于B方而言：

$$\nabla F(\omega)_B = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_B = \frac{1}{n}\sum_{i=1}^{n} d_i x_B$$

5. A、B双方求出梯度（密文）后，开始发送给arbiter方,然后它来解密梯度,再发送给A方和B方
6. A方、B方收到解密的梯度后，开始更新模型参数ω_A和 ω_B
7. 重复上述2~6过程，直到收敛或者达到指定的迭代次数

基于差分隐私(DP)技术

基于差分隐私的逻辑回归算法设计,其实整体思想和同态加密算法设计很相似,只不过在这里将加密技术改为添加噪声.主要流程如下:

1. B方选择小批量数据集X_t(t代表当前的迭代次数)，然后计算$IR_t^B = X_t^B \omega_t^B$

> 这里的IR就是intermediate result缩写,代表计算梯度过程中需要的中间结果。这里是因为线性模型的梯度是线性可分的

添加噪声，公式为：$Sec[IR_t^B] = IR_t^B + Z^B$ 。这样，$Sec[IR_t^B]$就是添加噪声后的IR_t^B。而后将此结果发送给A方

2. 这样，A方就获得了完整的$\omega x = \omega_A X_A + Sec[\omega_B X_B]$，只不过这里的$\omega_B X_B$添加了噪音。根据逻辑回归的整体损失函数，可以表示成(b代表的是当前的小批量)：

$$\nabla F(\omega) = \sum_{i=1}^{b}(\frac{y_i}{1+e^{-y_i\omega^T x_i}} - y_i)x_i = \sum_{i=1}^{b} d_i x_i$$

在这里将d_i表示成IR_t^A。然后IR_t^A添加噪声，$Sec[IR_t^A] = IR_t^A + Z_A$。将这个结果发送给B方

3. 在每次迭代经过一轮交互之后，A方、B方各自获得：

$$\nabla F(\omega)_A = \sum_{i=1}^{b} Sec[IR_t^A]x_A$$

$$\nabla F(\omega)_B = \sum_{i=1}^{b} Sec[IR_t^A]x_B$$

4. 下面的就和常规的逻辑回归一样了，A、B双方得到各自的梯度，于是更新子模型：ω_A和 ω_B
5. 重复1-4步骤便可。直到收敛或者达到指定的迭代次数

基于HE-DP技术

基于同态加密方案隐私泄露风险分析

以A方为例,当A方计算出密文梯度公式:

$$\nabla F(\omega)_A = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_A = \frac{1}{n}\sum_{i=1}^{n} d_i x_A$$

并且发送给arbiter方解密后,**A方解密后的梯度包含B的属性信息.很明显,梯度中含有数据信息.**(已有的论文提出了类似的观点),部分参考论文为

> 《exploiting unintended feature leakage in collaborative learning》
>
> 《model inversion attacks that exploit confidence information and basic countermeasures》

同样的,站在B的角度而言,**B方解密后的梯度包含A的属性和标签信息.很明显,梯度中含有数据信息.**

故同态加密方案仍存在隐私泄露的风险.

基于差分隐私方案的分析

B发送的数据:$Sec[IR_t^B] = IR_t^B + Z^B$.

因为这里A方需要获知B的每条数据信息,以便于后续计算梯度的中间结果d_i,所以B需要在**每条**数据上添加噪音.因此在计算全局敏感度的时候由于要考虑整体数据的变化范围,使得全局敏感度偏大,添加的噪声也较大,致使最终训练的模型预测精确度相较于其他方案有所降低.

同样的,A发送的数据:$Sec[IR_t^A] = IR_t^A + Z_A$也存在类似的全局敏感度较大的问题.

基于HE-DP的方案设计

我们的方案是综合考虑HE和DP的方案,在密文梯度上使用差分隐私(DP)技术扰动数据，保护用户数据信息.具体而言可以分为两部分:计算密文梯度,扰动数据.

计算密文梯度

这部分与同态加密的方案基本一致,最终A方\B方会得到密文梯度.分别如下:

$$\nabla F(\omega)_A = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_A = \frac{1}{n}\sum_{i=1}^{n} d_i x_A$$

$$\nabla F(\omega)_B = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_B = \frac{1}{n}\sum_{i=1}^{n} d_i x_B$$

扰动数据

对于A方而言,如果它直接从arbiter方接收到解密后的梯度($\nabla F(\omega)_A$),那么这个梯度实际上含有的是B方$\omega^T x_B$的部分数据信息.

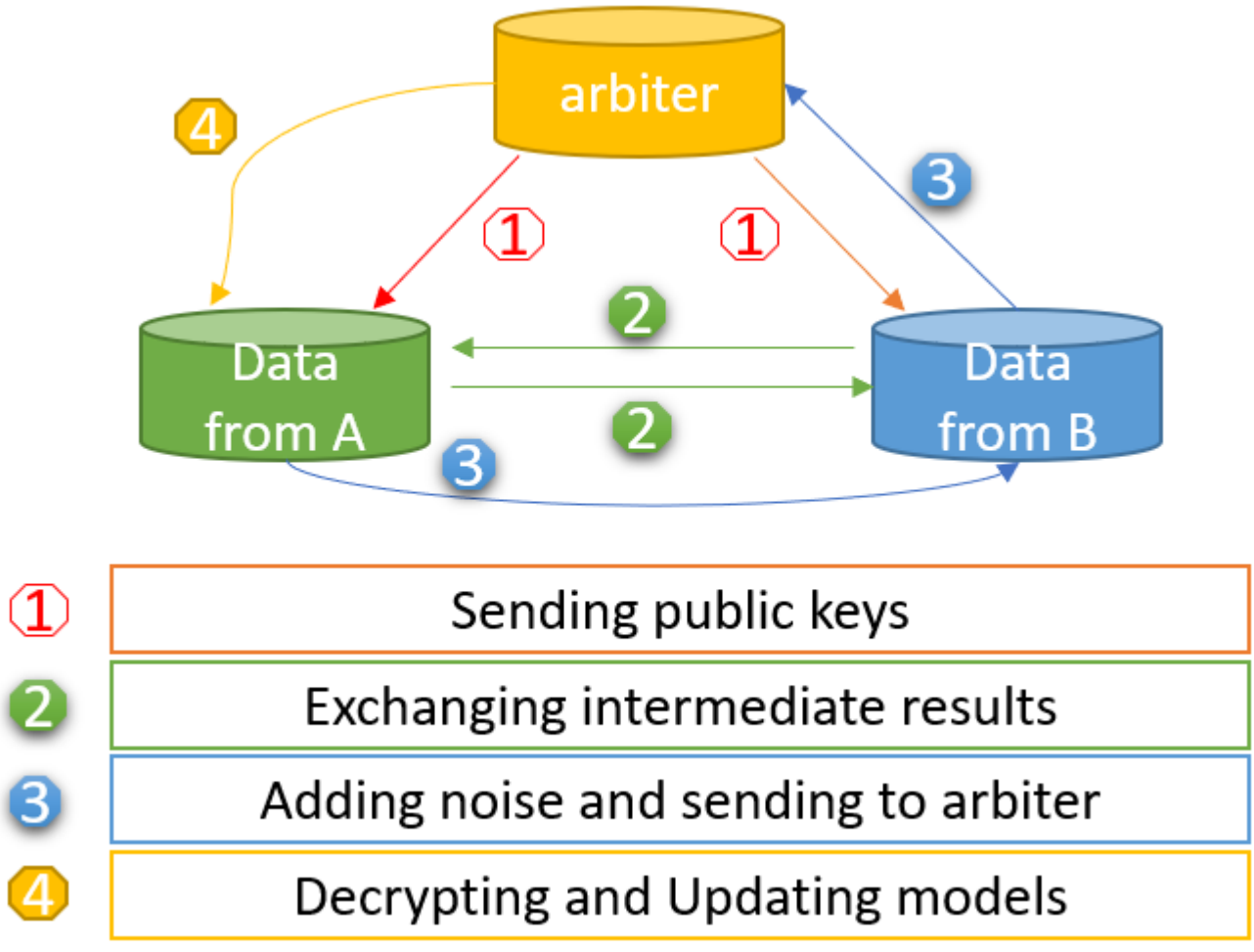
所以A方接收到的梯度($\nabla F(\omega)_A$)应当是经过B扰动后的梯度,又因为这个梯度是密文状态下,所以是由B方将$\nabla F(\omega)_A$扰动后发送给arbiter方,由arbiter解密后再发送给A方.

这个时候A方即使得到的是解密后的梯度,但是他所得到的梯度中B方的数据信息仍然是扰动后的数据信息.并不能从中推知B的数据隐私.

经过上述分析,我们会发现扰动数据阶段实际步骤分为以及几步:

- A方计算出密文梯度$\nabla F(\omega)_A$,然后发送给B方
- B方接收到A方的密文梯度$\nabla F(\omega)_A$,添加噪声扰动密文梯度.然后发给arbiter方

如下图:

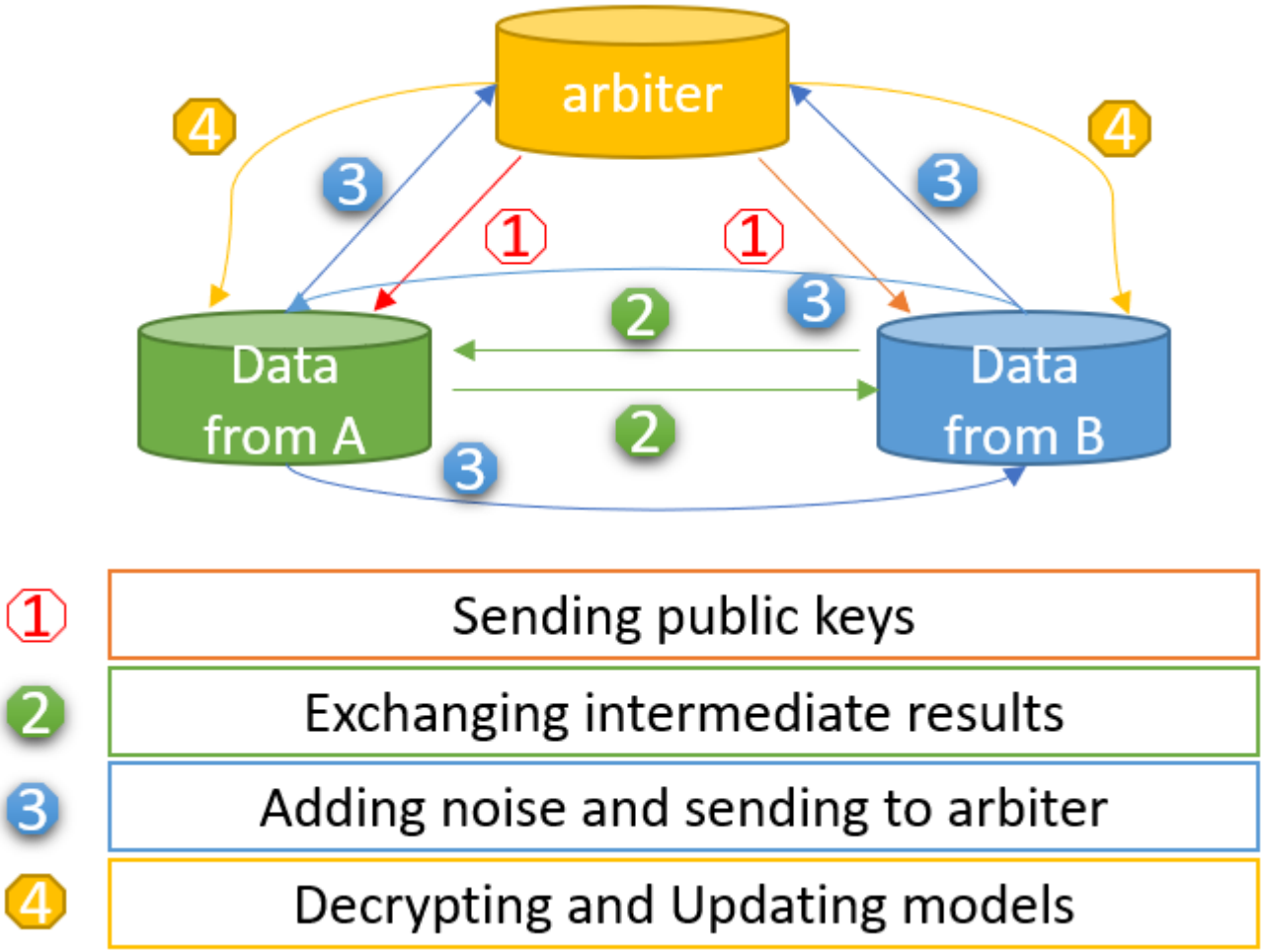


- arbiter解密梯度后,发送给A方
- 这样,A方所获得的便是扰动后的梯度.

同样的,站在B方角度而言.如果B方是攻击者,也会获得A方的数据隐私.故A方也会扰动数据,具体流程如下:

- B方计算出密文梯度$\nabla F(\omega)_B$,然后发送给A方
- A方接收到B方的密文梯度$\nabla F(\omega)_B$,添加噪声扰动密文梯度.然后发给arbiter方
- arbiter解密梯度后,发送给B方
- 这样,B方所获得的便是扰动后的梯度.

结构如下:



以上便是我们关于HE-DP方案设计

实验数据

在纵向联邦学习的线性模型中，由于模型交互逻辑具有相似性，所以这里以纵向逻辑回归为例来进行实验。

数据集

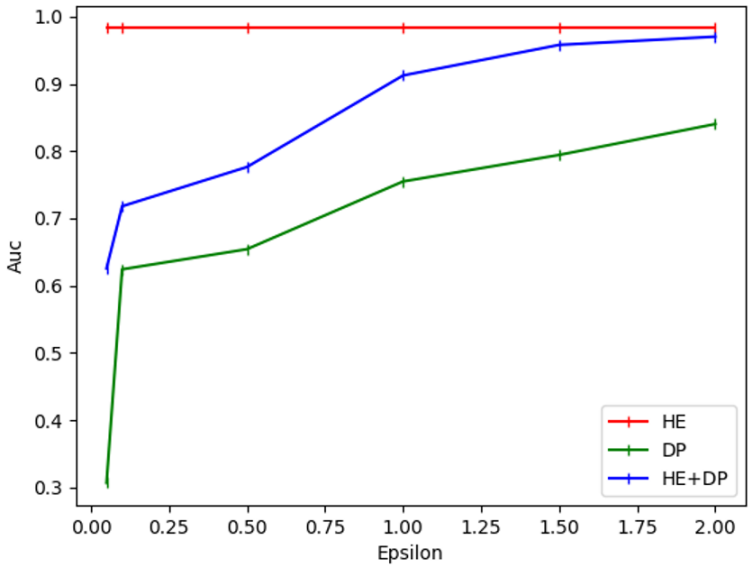
这里选用Kaggle和UCI Machine Learning Repository 中的不同数据和不同分布进行实验，这里仅以Breast数据集、Adult数据集为例，具体属性信息如图:

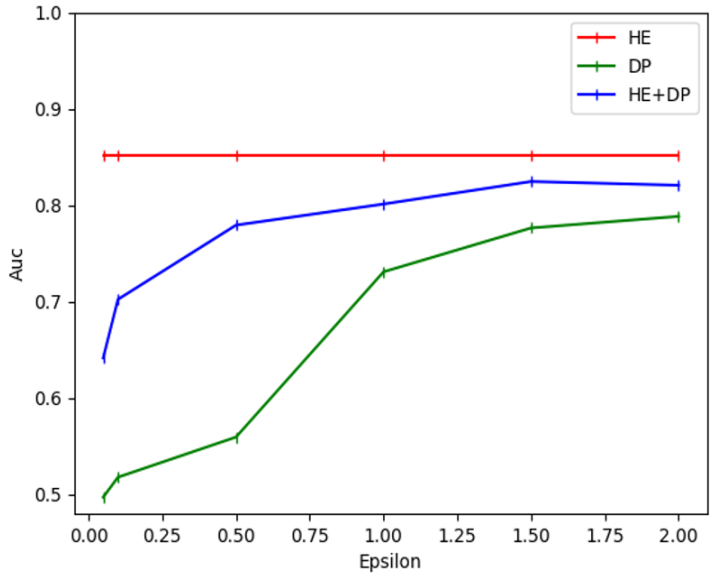
Datasets	Task	Attributes(guest)	Attributes(Host)
Breast	Binary Classification	11	20
Adult	Binary Classification	7	8

方法对比

我们这里采用三种方法，分别是基于同态加密HE的纵向逻辑回归，基于差分隐私DP的纵向逻辑回归，基于HE-DP的纵向逻辑回归来进行对比。

衡量指标是ROC曲线下的面积:AUC值去衡量分类器效果，实验效果如下：





其中左右两图分别代表的数据集分别是：Breast和Adult数据集

通过实验，我们发现随着隐私预算(ϵ)的增大，HE-DP方法的分类准确度高于DP方法。原因在于不同于DP方法直接在每条数据上添加噪音，HE-DP方法是在每一次迭代后的密文梯度上添加噪声扰动数据，所以全局敏感度较小，噪音规模也较小。

我们还发现由于HE曲线未采用噪声扰动数据，所以最终得到的模型分类精确度是三者中最高的，但是我们所采用的HE-DP方法，当隐私预算足够大时，模型分类效果仍接近HE方法。

结论：我们团队所设计的基于HE-DP的用于线性纵向联邦学习的方案，在隐私保护水平和精确度之间获取一个均衡，在保证降低用户隐私泄露风险的前提下，也能得到较好的训练结果。

同时用户可以灵活的调整隐私预算$epsilon$来根据自身的客观条件去决定需要更高的安全性还是更好的建模效果。

总结而言就是一句话：使用同态加密技术保证了计算过程的安全性，使用差分隐私技术保证了计算结果的安全性。同样，我们将逻辑回归延伸到了线性回归和泊松回归，具体的公式推理附件2.

由于纵向神经网络使用的安全技术主要是同态加密，所以我们的下一步的工作的安排就是研究并提升纵向神经网络的安全性。

附件1：主要参考论文

> 《Federated Machine Learning: Concept and Applications》
>
> 《Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption》
>
> 《hybrid differentially private federated learning on vertically partitioned data》
>
> 《Train faster, generalize better: Stability of stochastic gradient descent》

附件2：公式推导

1. 关于ΔIR_A与ΔIR_B的几何意义分析

A方和B方每次迭代的过程中，会计算出如下两个子梯度：

$$\nabla F(\omega)_A = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_A = \frac{1}{n}\sum_{i=1}^{n} d_i x_A$$

$$\nabla F(\omega)_B = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x - 0.5y)x_B = \frac{1}{n}\sum_{i=1}^{n} d_i x_B$$

δ：差分隐私参数，允许一定的误差，这里的δ一般取值很小。因为这里采用的高斯噪声

ϵ：隐私预算

$\beta_\theta : smooth\ parameters$

L：李普希兹常数

$e : epoches$，表示数据的完整的使用次数

T：迭代次数

$eta : learning\ rating$。学习率

$b : mini - batch\ size$

k：梯度剪切参数，防止梯度爆炸

$\beta_y : smooth\ parameters$

$k_y : target\ bound$

$length : host$方或者$guest$方的数据维度

其中ΔIR_A表示的就是A方属性和标签变化的最大敏感度(迭代了T次)

$$[\Delta([IR_t^A]_{t=1}^T)]^2$$
$$\leq 4\beta_\theta^2 L^2 \frac{e^2 T\eta^2}{b} + 8(\beta_\theta k + \beta_y k_y)\beta_\theta L\frac{e^2\eta}{b} + 4(\beta_\theta k + \beta_y k_y)^2 e$$

其中ΔIR_B表示的就是B方属性变化的最大敏感度(迭代了T次)

$$\Delta([IR_t^B]_{t=1}^T) \leq \sqrt{\frac{T(2e\eta L)}{b}^2 + \frac{8ke^2\eta L}{b} + 4ek^2}$$

2. 线性回归的公式推理

2.1 逻辑回归最大敏感度推理

A方扰动数据

$$[[\nabla F(\omega)_B]] = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x_i - 0.5y_i)x_B = \frac{1}{n}\sum_{i=1}^{n} d_i x_B$$

全局敏感度

$$\frac{1}{n}\Delta IR^A||dimen_B||_2$$

B方扰动数据

$$[[\nabla F(\omega)_A]] = \frac{1}{n}\sum_{i=1}^{n}(0.25\omega^T x_i - 0.5y_i)x_A = \frac{1}{n}\sum_{i=1}^{n} d_i x_A$$

全局敏感度

$$\frac{1}{n}0.25\Delta IR^B||dimen_A||_2$$

3. 线性回归最大敏感度推理

A方扰动数据

$$[[\nabla F(\omega)_B]] = (-2)\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\omega)x_B$$

全局敏感度

$$\frac{1}{n}\Delta IR^A||dimen_B||_2$$

3.2 B方扰动数据

$$[[\nabla F(\omega)_A]] = (-2)\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\omega)x_A$$

全局敏感度

$$\frac{1}{n}2\Delta IR^B||dimen_A||_2$$

4.泊松回归最大敏感度推理

A方扰动数据

$$[[\nabla F(\omega)_B]] = \frac{1}{n}\sum_{i=1}^{n}(y_i - e^{W^Tx_i})x_B$$

全局敏感度

$$\frac{1}{n}\Delta IR^A||dimen_B||_2$$

B方扰动数据

$$[[\nabla F(\omega)_A]] = \frac{1}{n}\sum_{i=1}^{n}(y_i - e^{W^Tx_i})x_A$$

全局敏感度

$$\frac{1}{n}[1-(dimen_A+dimen_B)k+\frac{1}{3}(dimen_A+dimen_B)^2k^2]\Delta IR^B||dimen_A||_2$$

如下为泊松回归的模型中最大敏感度的部分推理：

其中我们设定的条件是:属性A、B双方的1范数小于等于1；然后模型参数ω的1范数小于等于k

$$||x|| \le 1$$

$$||\omega|| \le k$$

假设A、B双方的属性维度是分别是：$dimen_A$和 $dimen_B$.为了便于推理公式，我们假设$\omega_A^Tx_A$设定为A方；假设$\omega_B^Tx_B$设定为B方；相邻数据集设定出的$\omega_B^{T'}x_B$设定为B'.我们使用泰勒公式将指数展开到第三项，详细结果是：

$$e^{\omega_A^Tx_A+\omega_B^Tx_B} - e^{\omega_A^Tx_A+\omega_B^{T'}x_B}$$

$$= 1+(\omega_A^Tx_A+\omega_B^Tx_B)+\frac{1}{2}(\omega_A^Tx_A+\omega_B^Tx_B)^2+\frac{1}{6}[\omega_A^Tx_A+\omega_B^Tx_B]^3$$
$$-1-\omega_A^Tx_A-\omega_B^{T'}x_B-\frac{1}{2}[\omega_A^Tx_A+\omega_B^{T'}x_B]^2-\frac{1}{6}[\omega_A^Tx_A+\omega_B^{T'}x_B]^3$$
$$=(1+\omega_A^Tx_A)\Delta IR^B+\frac{1}{2}[(\omega_B^Tx_B+\omega_B^{T'}x_B)\Delta IR^B]+$$
$$\frac{1}{6}[2\omega_A^T{x_A}^2+2\omega_A^Tx_A(\omega_B^Tx_B+\omega_B^{T'}x_B)+\omega_B^T{x_B}^2+\omega_B^{T'}{x_B}^2]\Delta IR^B$$
$$\le [1+(dimen_A+dimen_B)k+\frac{1}{3}[(dimen_A+dimen_B)^2k^2]]\Delta IR^B$$