王大星,周强,滕济凯.基于同态加密和牛顿迭代法的数据隐私保护模型[J].湖南科技大学学报(自然科学版),2024,39(2):69-74.doi:10.13582/j.cnki.1672-9102.2024.02.009

WANG D X, ZHOU Q, TENG J K. Data Privacy Protection Model Based on Homomorphic Encryption and Newton Iterative Method[J]. Journal of Hunan University of Science and Technology (Natural Science Edition),2024,39(2):69-74.doi:10.13582/j.cnki.1672-9102.2024.02.009

基于同态加密和牛顿迭代法的数据隐私保护模型

王大星[1*],周强[1],滕济凯[2]

(1.滁州学院 数学与金融学院,安徽 滁州 239000;2.青岛理工大学 理学院,山东 青岛 266555)

摘 要:为解决目前机器学习面临的严重隐私泄露问题,提出基于同态加密的大数据隐私保护 Logistic 回归模型.利用 Logistic 回归算法对加密数据进行训练和预测,确保在整个过程中不会泄露任何隐私,同时,采用 Paillier 同态加密算法加密训练数据,利用牛顿迭代算法建立适用于密文数据集的逻辑回归模型.分别在 MNIST 和 Dermatology 数据集上执行该模型算法,通过进一步解密明文之后进行评估,进而计算所提模型的准确率.最后,将所提出的模型与相关文献模型进行比较,结果表明:所提模型具有良好的性能和较高的准确率,输出结果与未加密明文运算的输出结果一致,且不影响模型的准确率.所提出的模型可以用于构建二进制分类模型的隐私保护和通过逻辑回归建模的各种问题.

关键词:大数据;回归模型;隐私保护;同态加密

中图分类号:TP309.7 **文献标志码**:A **文章编号**:1672-9102(2024)02-0069-06

Data Privacy Protection Model Based on Homomorphic Encryption and Newton Iterative Method

WANG Daxing[1], ZHOU Qiang[1], TENG Jikai[2]

(1. School of Mathematics and Finance, Chuzhou University, Chuzhou 239000, China;
2. College of Science, Qingdao Technological University, Qingdao 266555, China)

Abstract: In order to solve the problem of serious privacy leakage faced by machine learning, this paper proposes a Logistic regression model of big data privacy protection based on homomorphic encryption. The model uses Logistic regression algorithm to train and predict encrypted data, and the whole process will not reveal any privacy. The training data is encrypted by using Paillier homomorphic encryption algorithm, and the logistic regression model suitable for ciphertext data set is established by using Newton iterative algorithm. The algorithm can be used as a privacy preserving technique to construct binary classification model, and can be applied to various problems that can be modeled by logistic regression. This paper implements the algorithm on MNIST and dermatology data sets respectively, evaluates the model after further decrypting the plaintext, and then calculates the accuracy of the model. Finally, comparing the model of this paper with the conclusions of relevant literatures, it can be seen that the model of this paper has high accuracy, which shows the feasibility of the model in practical application.

Keywords: big data; regression model; privacy protection; homomorphic encryption

收稿日期:2022-10-28

基金项目:安徽省高等学校科研计划重大项目资助(2022AH040148)

* **通信作者**,E-mail: daxingwang@chzu.edu.cn

机器学习是人工智能中的一类方法,其特征是不提供特定问题的解决方案,而是通过学习为一系列相似问题找到解决方案的过程[1].机器学习理论出现于 20 世纪 60 年代初期,推动了技术复杂的学习系统理论和实践的发展.尽管机器学习的应用范围正在不断扩大,但随着机器学习的兴起,安全问题已成为该领域的研究热点.例如,许多医疗决策都依赖逻辑回归模型,而生物医学数据通常包含关于个人的机密信息[2].因此,数据的私密性和安全性是主要问题,特别是在部署外包分析工具时的安全问题尤为重要.

目前,基于密码学的安全计算的研究很多.多方计算技术也被应用于保护隐私的逻辑回归[3],但是,当一方不诚实地进行操作时,该方法很容易受到攻击,并且秘密共享的假设与外包计算的假设完全不同.同态加密是一种加密系统,可以对加密的数据执行某些算术运算,与以明文形式执行的运算结果相对应.WU 等[4]使用 Paillier 密码系统[5]和多项式逼近 Logistic 函数,发现随着逼近多项式次数的增加,计算成本也在呈指数增长;XIE 等[6]使用加法同态方案来汇总一些中间统计数据,但该方法需要依靠客户端解密这些中间统计数据,需要消耗大量的计算成本;CRAWFORD 等[7]的方法在训练小型 Logistic 回归模型方面有良好的表现,但是只针对计算特征非常少的数据.目前,很多的研究是基于 KIM 等[8]的工作,也使用基于同态加密的机器学习,然而,加密数据的大小和学习时间高度依赖于特征的数量,因此,对于大型数据集来说,性能在存储和计算成本方面并不实用.

Logistic 回归是机器学习中用于解决二进制分类问题的一种流行技术,它从训练阶段开始,根据先前收集的预测变量(称为协变量)值和相应结果进行预测.训练阶段之后是评估模型准确性的测试阶段,通过评估协变量模型输出的参数与已知结果进行比较来完成验证.当模型的分类等于大多数测试数据的结果时,模型被认为具有应用价值.Logistic 回归提供了一种简单而有效的方法来解决多种问题.在医学中,Logistic 回归用于根据观察到的患者特征预测发展某种疾病的风险;在政治中,它用于根据个人数据如年龄、收入、性别和以往的投票等预测选举结果;在金融领域,Logistic 回归用于预测房主拖欠抵押贷款或信用卡交易的可能性.BOS 等[9-10]考虑借助同态加密的私有逻辑回归,但都假定逻辑模型已经过训练并且可以公开获得;AONO 等[11]通过同态加密探索安全的逻辑回归模型,然而,该模型将具有挑战性的同态计算转移到可信数据源和可信客户端,使得该方法只能使用加法同态加密方案完成训练过程中比较简单的部分.国内学者也探索了通过同态加密进行的安全逻辑回归,但是,这些方案将难以同态执行的计算转移到受信任的数据源和受信任的客户端[12-16].本文根据文献[17]中的牛顿迭代法构造一种新的同态加密回归算法,用于训练适用于同态加密数据集的逻辑回归模型.该算法可以用于构建二进制分类模型的隐私保护技术和通过逻辑回归建模的各种问题.试验结果表明:该模型具有良好的性能,输出的结果与未加密明文的运算输出结果一致,且不影响模型的准确率.

1 Logistic 回归模型

Logistic 回归是一种机器学习算法,用于学习分类模型.为了简化介绍,本文将重点放在二进制分类上.假设获取的数据信息为 (x_i,y_i),其中 $x_i \in \mathbf{R}^d$,$y_i \in \{0,1\}$.在给定 x 的情况下,目的是预测 y 的值,考虑以下模型:

$$\Pr[y=1 \mid x] = \frac{1}{1+\mathrm{e}^{(-w_0-\sum_{i=1}^{n}x_iw_i)}} = \frac{1}{1+\mathrm{e}^{-\boldsymbol{x}_i^{'\mathrm{T}}\boldsymbol{w}}}. \tag{1}$$

式中:w_0 为回归系数初值;w_i 为回归系数;n 为回归系数的维数;$\boldsymbol{w}$ 为需要寻找的长度为 $d+1$ 的权重向量;$\boldsymbol{x}_i^{'} = (1,x_i) \in \mathbf{R}^{d+1}$.

给定一组训练数据 $\{(y_i,\boldsymbol{x}_i)\}_{i=1}^{n}$,模型的目的是寻找最佳的权重向量 $\boldsymbol{w}^*$.由于 $1-\frac{1}{1+\mathrm{e}^{-z}} = \frac{1}{1+\mathrm{e}^{z}}$,那么

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{\operatorname{argmax}}\left\{\prod_{y_i=1}\frac{1}{1+\mathrm{e}^{(-\boldsymbol{x}_i^{'\mathrm{T}}\boldsymbol{w})}}\cdot\prod_{y_i=0}\frac{1}{1+\mathrm{e}^{(\boldsymbol{x}_i^{'\mathrm{T}}\boldsymbol{w})}}\right\} =$$

$$\underset{\boldsymbol{w}}{\operatorname{argmax}}\left\{\prod_{i=1}\frac{1}{1+\mathrm{e}^{(-z_i^{\mathrm{T}}\boldsymbol{w})}}\right\} = \underset{\boldsymbol{w}}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}\log(1+\mathrm{e}^{(-z_i^{\mathrm{T}}\boldsymbol{w})})\right\}. \tag{2}$$

式中:$\boldsymbol{w}^*$ 为最佳权重向量;$\boldsymbol{z}_i = y_i^{'}\cdot\boldsymbol{x}_i^{'}$;$y_i^{'} = 2y_i-1 \in \{-1,1\}$.

给定的训练集定义损失函数:

$$J(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n}\log(1 + \mathrm{e}^{-z_i^{\mathrm{T}}\boldsymbol{w}}). \tag{3}$$

那么该优化问题实际上就是求解合适的权重向量 $\boldsymbol{w}$,使得 $J(\boldsymbol{w})$ 最小.梯度下降法是一种通过沿梯度移动找到函数的局部极值的方法.计算损失函数 $J(\boldsymbol{w})$ 关于自变量 $\boldsymbol{w}$ 的梯度得到:

$$\nabla_{\boldsymbol{w}}J(\boldsymbol{w}) = -\frac{1}{n}\sum_{i=1}^{n}\sigma(-z_i^{\mathrm{T}}\boldsymbol{w})\cdot z_i. \tag{4}$$

式中:$\sigma(x) = \dfrac{1}{1 + \mathrm{e}^{-x}}$.

输入初始化权重 $w^{(0)}$,记 α_t 为第 t 步的学习速率,梯度下降法在第 t 步的更新回归参数如式(5)所示.

$$w^{(t+1)} = w^{(t)} + \frac{\alpha_t}{n}\sum_{i=1}^{n}\sigma(-z_i^{\mathrm{T}}\boldsymbol{\beta}^{(t)})\cdot z_i. \tag{5}$$

式中:$\boldsymbol{\beta}^{(t)}$ 为第 t 步得到的最佳权重向量.

2　同态加密理论

数据的隐私可以通过加密算法来实现.同态加密允许用户无需解密明文而直接对密文进行操作,却能得到和对明文直接进行操作一样的效果.同态加密过程如图 1 所示.如图 1 所示,Alice 拥有隐私明文 m,将明文加密为 $E(m)$ 发送给 Bob.Bob 对密文进行某种计算或变换得到 $E[f(m)]$ 并发送给 Alice,Alice 解密得到 $f(m)$. 在整个过程中 Bob 对明文 m 的内容一无所知,从而为 Alice 实现了有效的隐私保护.

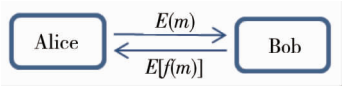

图 1　同态加密

Paillier 同态加密算法的简要表述如下:

1)密钥生成:随机选择 2 个长度相等的大素数 p,q,计算 $n = pq, \varphi(n) = (p-1)(q-1)$,选择随机数 $g \in \mathbf{Z}_{n^2}^*$,则公钥 $\mathrm{pk} = (n,g)$,私钥 $\mathrm{sk} = [\varphi(n), \varphi(n)^{-1} \bmod n]$.

2)加密:输入明文 m,选择 $r \in \mathbf{Z}_{n^2}^*$,并且 $0 < r < n$,输出密文 $c = g^m \times r^n \bmod n^2$.

3)解密:输入密文 c,输出明文 $m = L(c^{\varphi(n)} \bmod n^2) \times \varphi^{-1}(n) \bmod n$,此处函数的定义为 $L(x) = \dfrac{x-1}{n}$.

3　同态逻辑回归算法

3.1　牛顿迭代法

牛顿迭代法又称牛顿-拉夫逊方法(Newton-Raphson Method),它是牛顿在 17 世纪提出的一种在实数域和复数域上近似求解方程的方法.运用牛顿迭代法求解式(1)中的权重向量 $\boldsymbol{w}$ 的值,用于计算单变量函数 $f(x)$ 根的 Newton-Raphson 方法的迭代公式为

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_{k+1})}. \tag{6}$$

式中:x_{k+1} 为 $f(x)$ 根的第 $k+1$ 次近似值;x_k 为 $f(x)$ 根的第 k 次近似值;$f(x_k)$ 为 x_k 处的函数值;$f'(x_{k+1})$ 为 x_{k+1} 处的导数值.

计算的目标函数是多元函数 $J(\boldsymbol{w})$,因此,参数向量 $\boldsymbol{w}$ 的第 $k+1$ 次迭代为

$$w_{k+1} = w_k - \boldsymbol{H}^{-1}(w_k)\nabla_{\boldsymbol{w}}J(w_k). \tag{7}$$

式中:$\boldsymbol{H}(\boldsymbol{w}) = \nabla_{\boldsymbol{w}}^2 J(\boldsymbol{w})$ 为 $J(\boldsymbol{w})$ 的 Hessian 矩阵,也即函数 $J(\boldsymbol{w})$ 对其分量 w_k 的二阶偏导数矩阵,$\boldsymbol{H}_{i,j} = \dfrac{\partial^2 J}{\partial w_i \partial w_j}$,因此 $\boldsymbol{H}(\boldsymbol{w}) = -\sum_{i=1}^{d+1}[1 - \sigma(y_i \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i)]\cdot\sigma(y_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i)(y_i\boldsymbol{x}_i)^2$.

3.2　同态加密 Logistic 回归

首先,描述基本的逻辑回归训练过程,然后运用牛顿迭代法进行优化.考虑训练时输入大数据的密文包,本文采用的更新每次迭代的权重参数的公式为

$$w_{i+1} = w_i - \gamma \cdot \nabla_w J(w_i). \tag{8}$$

式中:γ 为步长.

由于非多项式函数在执行同态加密算法时的计算成本太大,因此,必须选择合适的多项式函数近似逻辑回归模型中的激活函数 $\sigma(x) = \dfrac{1}{1 + e^{-x}}$. 本文给出如下备选函数:

1)在 $x = 0$ 处的泰勒多项式:$y_1 = 0.5 + 0.25x - 0.021x^3$.

2)最小二乘法拟合多项式:$y_2 = 0.5 + 0.15x - 0.001\ 5x^3$;$y_3 = 0.5 + 0.22x - 0.008x^3 + 0.000\ 17x^5$.

图 2 是标准函数及 3 种激活函数的图像对比.由图 2 可知:泰勒展开式提供了在某个点附近的近似曲线,而多项式拟合函数则在一个区间范围内有良好的近似.因此,本文选择的激活函数是多项式函数 y_3.

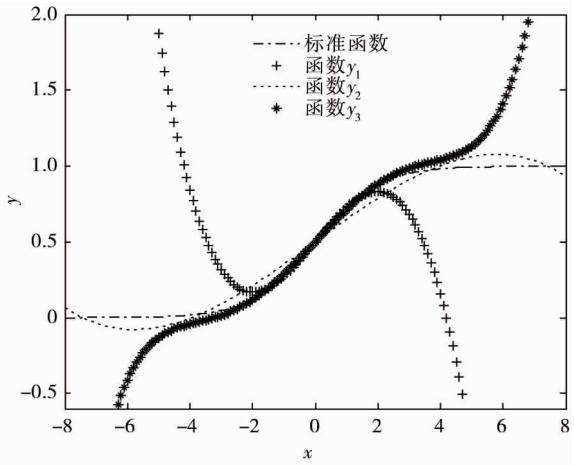

图 2 激活函数

考虑牛顿迭代法在求解 Hessian 矩阵的逆运算时的计算开销很大,而使用同态加密以隐私保护的方式估计 Logistic 回归模型的参数将进一步增加计算的难度.因此,所采用的方法是修改 Hessian 矩阵,以使其有可能在加密域中有效地进行计算.根据文献[17]中所讨论的方法,每次迭代都在更新的 Hessian 矩阵 $\boldsymbol{H}(\boldsymbol{w}) = \nabla_{\boldsymbol{w}}^2 J(\boldsymbol{w})$ 可以用固定的矩阵 $\boldsymbol{W}$ 代替,而不影响牛顿迭代法的收敛性.该方法的前提是 $\boldsymbol{H}(\boldsymbol{w}) - \boldsymbol{W}$ 是非负定的.那么,给定矩阵 $\boldsymbol{H}$,牛顿迭代公式可以简化为

$$w_{k+1} = w_k - \boldsymbol{H}^{-1} \nabla_{\boldsymbol{w}} J(w_k). \tag{9}$$

在本文算法中,取定 Hessian 矩阵为 $\boldsymbol{W} = -\dfrac{1}{4}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. 因此,只要给定输入的训练数据值的范围,并且选择牛顿迭代法的初始值,考虑训练数据的维数,就可以估计求逆运算的代价.具体算法流程描述如下:

密文 Logistic 回归算法

Input:训练数据集 $\boldsymbol{X}(n, d+1)$,标签矩阵 $\boldsymbol{Y}(n, 1)$,牛顿迭代初值 a_0,迭代次数 t.

%算法中的矩阵或数据集是以 MATLAB 程序代码格式表示的,如 $\boldsymbol{X}(i, j)$ 表示 i 行 j 列的矩阵.

Output:回归参数 $\boldsymbol{w}$.

```
w = 0.001 * ones(d+1,1); % ones(i, j)表示 i 行 j 列的全 1 矩阵.
sum = zeros(n,1);
for i = 1 to n do
    for j = 1 to d+1 do
        sum(i) = sum(i) + X(i, j)
    end
end
for j = 1 to d+1 do
    temp = 0;
    for i = 1 to n do
        temp = temp + X(i, j) * sum(i)
    end
    W(i, j) = (-1/4) * temp;
    W^-1(i, j) = 2a0 - W(i, j) * a0^2;
end
for k = 1 to t do
    for i = 1 to n do
        c = c + [(1/2) + (1/4) * Y(i) * X(i, :) * w] * Y(i) * X(i,:); % X(i, :)表示矩阵 X 的第 i 行.
    end
    w = w - W^-1 * c;
end
```

由于该模型输入的训练数据是同态加密的密文,所以,在整个训练过程或者预测过程中,对于该模型的任何使用者来说数据都是保密的.计算中间结果都是在密文下进行,敌手即使截获信道消息也无法解密.使用算法的双方之间传输的明文是加上掩码的模型参数梯度,由于敌手不知道掩码,无法获得关于梯度的信息.因此,敌手为第三方时,也无法通过信道获取任何敏感信息.

4　试验与分析

本文算法的实现基于同态加密开源库 HElib,同态加密的密钥长度为 1 024 bit.所有试验都在 Intel i5-7400 CPU 3 GHz 环境下,在 MNIST 数据集上完成.

MNIST 是一个著名的手写体数字灰度图像数据集,在很多资料中,这个数据集都会被作为深度学习的入门样例.MNIST 数据集主要由一些手写数字的图片和相应的标签组成,图片一共有 10 类,分别对应 0~9共 10 个数字.MNIST 数据集中的每一张图片都代表 0~9 中的一个数字,每张图片包含 28×28 像素的灰度图像,各个像素的取值在 0~255.每个图像数据都相应地标有数字标签.文献[16]中的 Dermatology 数据集研究的是某类皮肤疾病的鉴别诊断,该数据集包含 358 个样本,每个样本包含 34 个特征.在 MNIST 数据集和 Dermatology 数据集上,将本文算法所得结果与文献[8]和文献[16]的各指标进行比较,结果如表 1 所示.

表 1　MNIST 和 Dermatology 数据集上的实现结果

模型	数据集	样本数	特征数	迭代次数	加密方案	训练时间/min	准确率/%
文献[8]	MNIST	1 000	100	33	Paillier	64	84.60
文献[16]		1 000	100	21	HEAAN	77	76.40
本文算法		1 000	100	30	Paillier	50	87.00
文献[8]	Dermatology	358	34	11	Paillier	24	86.12
文献[16]		358	34	7	HEAAN	37	77.18
本文算法		358	34	9	Paillier	20	90.00

由表 1 可知:本文算法在 MNIST 数据集上执行 30 次迭代的时间为 50 min,平均每次约 1.67 min,而 Dermatology 数据集上只执行了 9 次迭代,平均每次约 2.22 min.通过进一步解密明文之后评估模型,得到本文模型在 MNIST 数据集和 Dermatology 数据集上的准确率分别为 87%和 90%.将本文的模型与文献[8]和文献[16]的模型进行比较可知:本文模型具有较高的准确率,表现了在现实应用中的可行性.

进一步对 MNIST 明文数据集进行逻辑回归算法的实现,并将所得结果与密文学习的结果进行比较,以评估同态加密中的噪声对机器学习表现的影响(如收敛速度和准确性),明文和密文回归准确率的对比结果如图 3 所示.由图 3 可知:在训练过程的早期,密文训练每次迭代的准确性与明文略有不同,但最终收敛于最后一步.

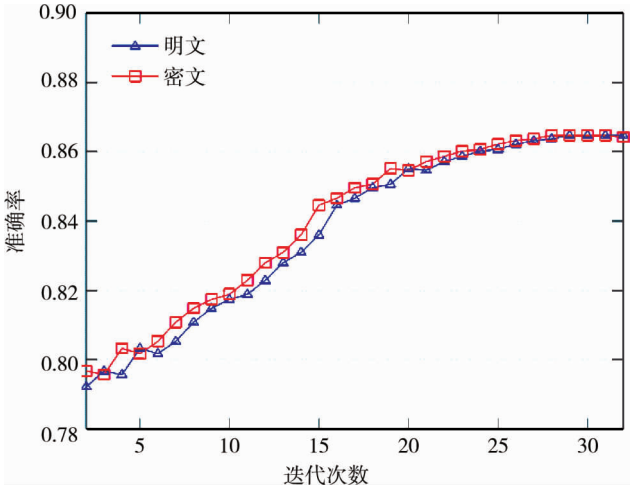

图 3　明文和密文回归准确率对比

5 结论

1)所提出的迭代方法可以对同态加密的输入数据进行逻辑回归训练,不仅可以实现训练过程的隐私保护,而且在训练过程中,协作的双方不能获得对方的训练数据及其模型参数信息.

2)在预测过程中,保护访问者的私有数据不会泄露给部署模型的服务器.所提出的方法可用于以隐私保护的方式将逻辑回归的训练阶段外包给云服务.

3)所提方法在 MNIST 和 Dermatology 数据集上的准确性比训练明文的逻辑回归模型标准算法稍高,表现出在对大型加密数据进行逻辑回归训练的可行性.同时,所提方法也可以很容易地应用于其他机器学习算法如神经网络等,这些将会作为后续的研究工作.

参考文献:

[1] YANG Q, LIU Y, CHEN T J, et al. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 12.

[2] HAN K, HONG K, CHEON J H, et al. Logistic regression model training based on the approximate homomorphic encrypted data at scale[C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, Hawaii, USA, 2019: 9466-9471.

[3] MOHASSEL P, ZHANG YP. SecureML: a system for scalable privacy-preserving machine learning[C]//2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017: 19-38.

[4] WU S, TERUYA T, KAWAMOTO J, et al. Privacy-preservation for stochastic gradient descent application to secure logistic regression[C]//The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013: 1-4.

[5] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[C]//International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg. 1999, 223-238.

[6] XIE W, WANG Y, BOKER S M, et al. PrivLogit: efficient privacy-preserving logistic regression by tailoring numerical optimizers[EB/OL]. 2016: arXiv: 1611.01170. http://arxiv.org/abs/1611.01170

[7] CRAWFORD J L H, GENTRY C, HALEVI S, et al. Doing real work with FHE: the case of logistic regression[C]// Proceedings of the 6th Workshop on Encrypted Computing & Applied Homomorphic Cryptography-WAHC'18. ACM, 2018: 1-12.

[8] KIM M, SONG Y, WANG S, et al. Secure logistic regression based on homomorphic encryption: design and evaluation[J]. JMIR Medical Informatics, 2018, 6(2): e19.

[9] BOS J W, LAUTER K, NAEHRIG M. Private predictive analysis on encrypted medical data[J]. Journal of Biomedical Informatics, 2014,50:234-243.

[10] DAVID B, DOWSLEY R, KATTI R, et al. Efficient unconditionally secure comparison and privacy preserving machine learning classification protocols[C]//Proceedings of the 9th International Conference on Provable Security. ACM, 2015, 9451: 354-367.

[11] AONO Y, HAYASHI T, PHONG L T, et al. Scalable and secure logistic regression via homomorphic encryption[C]// Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy. ACM, 2016: 142-144.

[12] 邹鸿珍.基于差分隐私的回归分析算法研究[D].赣州:江西理工大学,2016.

[13] 唐春明,魏伟明.基于安全两方计算的具有隐私性的回归算法[J].信息网络安全,2018(10):10-16.

[14] 马泽辉.基于逻辑回归算法的 Webshell 检测方法研究[J].信息安全研究,2019,5(4):298-302.

[15] 宋蕾,马春光,段广晗,等.基于数据纵向分布的隐私保护逻辑回归[J].计算机研究与发展,2019,56(10):2243-2249.

[16] 许心炜,蔡斌,向宏,等.基于同态加密的多分类 Logistic 回归模型[J].密码学报,2020,7(2):179-186.

[17] CHEN Y R, REZAPOUR A, TZENG W G. Privacy-preserving ridge regression on distributeddata[J]. Information Sciences, 2018, 451/452: 34-49.