

CLAIMS

What is claimed is:

1. A computer-implemented method for training multi-party secure logistic regression models, comprising:

dividing, by one or more processors, sample training data for a secure logistic regression model (SLRM) into a plurality of shares using secret sharing (SS) comprising hiding original data by encrypting the original data using random numbers generated by a plurality of secure computation nodes (SCN), such that the original data cannot be recovered by using a single share of the plurality of shares, wherein the sample training data is originating from the plurality of secure computation nodes configured to both produce and consume training data;

transmitting, by the one or more processors, each share to one of the plurality of secure computation nodes;

generating, by the one or more processors, updated parameters by iteratively updating current parameters associated with the SLRM using each share of the sample training data, wherein the iterative updating comprises:

for each SCN, calculating a value of A representing a column vector, wherein the calculating is performed based on the following equation:

$$A = X\Theta = X^L\Theta^L + X^R\Theta^R$$

wherein: $X^L\Theta^L$ and $X^R\Theta^R$ each represents a share of A for a SCN,

wherein calculating $X^L\Theta^L$ and $X^R\Theta^R$ comprises hiding original data provided from each SCN by using random numbers, and exchanging the hided data between the SCNs;

a predetermined iteration condition is satisfied; and

after iteratively updating the parameters associated with the SLRM, outputting a training result configured for use by each SCN of the plurality of secure computation nodes.

2. The computer-implemented method of claim 1, wherein an objective function of the SLRM is a logistic regression function that is optimized by using a mini batch Stochastic Gradient Descent (SGD) method, and wherein the logistic regression function is expressed as:

$$\theta[j] := \theta[j] - \frac{\alpha}{m} \sum_{i=1}^m (\text{sigmoid}(X_i^L \cdot \theta^L + X_i^R \cdot \theta^R) - Y_i) X_i[j] - \frac{\lambda}{m} \theta[j]$$

wherein:

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents a sample size of the mini batch SGD;

X represents a $m \times k$ sample matrix;

each row of the matrix X represents a sample;

X_i represents an i^{th} row of the matrix X ;

$[j]$ represents a j^{th} element of vector X_i ;

X^L represents a share of the sample data X that belongs to a respective SCN;

X^R represents a share of the sample data X that belongs to the respective SCN; and

θ represents a parameter column vector that is vertically dividable into θ^L and θ^R .

3. The computer-implemented method of claim 1, wherein iteratively updating parameters associated with the SLRM is performed according to an event-driven model.

4. The computer-implemented method of claim 1, wherein iteratively updating parameters associated with the SLRM comprises:

determining a column vector of a prediction error based on the following equation:

$$E = g(A) - Y_i;$$

wherein:

g represents a fitting function of a sigmoid function, the fitting function comprising a polynomial fitting function; and

determining a updated column vector based on equation:

$$\theta := (1 - \frac{\lambda}{m})\theta - \frac{\alpha}{m}X^T E$$

wherein:

$$X^T E = (X^L)^T E || (X^R)^T E,$$

wherein:

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents a sample size of a mini batch Stochastic Gradient Descent (SGD) method;

X represents a $m \times k$ sample matrix;

each row of the matrix X represents a sample;

X_i represents an i^{th} row of the matrix X ;

$[j]$ represents a j^{th} element of vector X_i ;

X^L represents a first share of the sample data X that belongs to a first SCN of the plurality of secure computation nodes;

X^R represents the second share of the sample data X that belongs to a second SCN of the plurality of secure computation nodes;

Y represents a prediction value;

Y_i represents a label of the i^{th} sample of the mini batch SGD method;

E represents an error vector;

R , L , and T represent indexes of the shares of a respective SCN; and

Θ represents a parameter column vector, and the column vector that is vertically dividable into Θ^L and Θ^R .

5. The computer-implemented method of claim 4, wherein calculating a share of A representing a column vector comprises:

encrypting original data provided by each SCN by using random numbers, and
exchanging the encrypted data between the plurality of secure computation nodes.

6. The computer-implemented method of claim 5, wherein each of the random numbers is at least one of a number, a vector, or a matrix.

7. The computer-implemented method of claim 6, wherein the predetermined iteration condition is an occurrence of a difference between two consecutive iteration results that is less than a predetermined threshold.

8. A non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising:

dividing sample training data for a secure logistic regression model (SLRM) into a plurality of shares using secret sharing (SS) comprising hiding original data by encrypting the original data using random numbers generated by a plurality of secure computation nodes, such that the original data cannot be recovered by using a single share of the plurality of shares, wherein the sample training data is originating from the plurality of secure computation nodes configured to both produce and consume training data;

transmitting each share to one of the plurality of secure computation nodes;

generating updated parameters by iteratively updating current parameters associated with the SLRM using each share of the sample training data, wherein the iterative updating comprises:

for each SCN, calculating a value of A representing a column vector, wherein the calculating is performed based on the following equation:

$$A = X\Theta = X^L\Theta^L + X^R\Theta^R$$

wherein: $X^L\Theta^L$ and $X^R\Theta^R$ each represents a share of A for a SCN,

wherein calculating $X^L\Theta^L$ and $X^R\Theta^R$ comprises hiding original data provided from each SCN by using random numbers, and exchanging the hidden data between the SCNs;

determining that a predetermined iteration condition is satisfied; and

after iteratively updating the parameters associated with the SLRM, outputting a training result configured for use by each SCN of the plurality of secure computation nodes.

9. The non-transitory, computer-readable medium of claim 8, wherein an objective function of the SLRM is a logistic regression function that is optimized by using a mini batch Stochastic Gradient Descent (SGD) method, and wherein the logistic regression function is expressed as:

$$\theta[j] := \theta[j] - \frac{\alpha}{m} \sum_{i=1}^m (\text{sigmoid}(X_i^L \cdot \theta^L + X_i^R \cdot \theta^R) - Y_i) X_i[j] - \frac{\lambda}{m} \theta[j]$$

wherein:

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents a sample size of the mini batch SGD;

X represents a m*k sample matrix;

each row of the matrix X represents a sample;

X_i represents an i^{th} row of the matrix X;

[j] represents a j^{th} element of vector X_i ;

X^L represents a share of the sample data X that belongs to a respective SCN;

X^R represents a share of the sample data X that belongs to the respective SCN; and

θ represents a parameter column vector that is vertically dividable into θ^L and θ^R .

10. The non-transitory, computer-readable medium of claim 8, wherein iteratively updating parameters associated with the SLRM is performed according to an event-driven model.

11. The non-transitory, computer-readable medium of claim 8, wherein iteratively updating parameters associated with the SLRM comprises:

determining a column vector of a prediction error based on the following equation:

$$E = g(A) - Y_i;$$

wherein:

g represents a fitting function of a sigmoid function, the fitting function comprising a polynomial fitting function; and

determining a updated column vector based on equation:

$$\theta := (1 - \frac{\lambda}{m})\theta - \frac{\alpha}{m}X^T E$$

wherein:

$$X^T E = (X^L)^T E || (X^R)^T E,$$

wherein:

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents a sample size of a mini batch Stochastic Gradient Descent (SGD) method;

X represents a $m \times k$ sample matrix;

each row of the matrix X represents a sample;

X_i represents the i^{th} row of the matrix X ;

$[j]$ represents the j^{th} element of vector X_i ;

X^L represents a first share of the sample data X that belongs to a first SCN of the plurality of secure computation nodes;

X^R represents the second share of the sample data X that belongs to a second SCN of the plurality of secure computation nodes;

Y represents a prediction value;

Y_i represents a label of the i^{th} sample of the mini batch SGD method;

E represents an error vector;

R , L , and T represent indexes of the shares of a respective SCN; and

Θ represents a parameter column vector, and the column vector Θ that is vertically dividable into Θ^L and Θ^R .

12. The non-transitory, computer-readable medium of claim 11, wherein calculating a share of A representing a column vector comprises:

encrypting original data provided by each SCN by using random numbers, and

exchanging the encrypted data between the plurality of secure computation

nodes.

13. The non-transitory, computer-readable medium of claim 12, wherein each of the random numbers is at least one of a number, a vector, or a matrix.

14. The non-transitory, computer-readable medium of claim 13, wherein the predetermined iteration condition is an occurrence of a difference between two consecutive iteration results that is less than a predetermined threshold.

15. A system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

dividing sample training data for a secure logistic regression model (SLRM) into a plurality of shares using secret sharing (SS) comprising hiding original data by encrypting the original data using random numbers generated by a plurality of secure computation nodes, such that the original data cannot be recovered by using a single share of the plurality of shares, wherein the sample training data is originating from the plurality of secure computation nodes configured to both produce and consume training data;

transmitting each share to one of the plurality of secure computation nodes;

generating updated parameters by iteratively updating current parameters associated with the SLRM using each share of the sample training data, wherein the iterative updating comprises:

for each SCN, calculating a value of A representing a column vector, wherein the calculating is performed based on the following equation:

$$A = X\Theta = X^L\Theta^L + X^R\Theta^R$$

wherein: $X^L\Theta^L$ and $X^R\Theta^R$ each represents a share of A for a SCN,

wherein calculating $X^L\theta^L$ and $X^R\theta^R$ comprises hiding original data provided from each SCN by using random numbers, and exchanging the hidden data between the SCNs;

determining that a predetermined iteration condition is satisfied; and

after iteratively updating the parameters associated with the SLRM, outputting a training result configured for use by each SCN of the plurality of secure computation nodes.

16. The system of claim 15, wherein an objective function of the SLRM is a logistic regression function that is optimized by using a mini batch Stochastic Gradient Descent (SGD) method, and wherein the logistic regression function is expressed as:

$$\theta[j] := \theta[j] - \frac{\alpha}{m} \sum_{i=1}^m (\text{sigmoid}(X_i^L \cdot \theta^L + X_i^R \cdot \theta^R) - Y_i) X_i[j] - \frac{\lambda}{m} \theta[j]$$

wherein:

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents the sample size of the mini batch SGD;

X represents a $m \times k$ sample matrix;

each row of the matrix X represents a sample;

X_i represents an i^{th} row of the matrix X ;

$[j]$ represents a j^{th} element of vector X_i ;

X^L represents a share of the sample data X that belongs to a respective SCN;

X^R represents a share of the sample data X that belongs to the respective SCN; and

θ represents a parameter column vector that is vertically dividable into θ^L and θ^R .

17. The system of claim 15, wherein iteratively updating parameters associated with the SLRM is performed according to an event-driven model.

18. The system of claim 15, wherein iteratively updating parameters associated with the SLRM comprises:

determining a column vector of a prediction error based on the following equation:

$$E = g(A) - Y_i;$$

wherein:

g represents a fitting function of a sigmoid function, the fitting function comprising a polynomial fitting function; and

determining a updated column vector based on equation:

$$\theta := (1 - \frac{\lambda}{m})\theta - \frac{\alpha}{m} X^T E$$

wherein:

$$X^T E = (X^L)^T E || (X^R)^T E, \text{ wherein:}$$

α and λ comprising hyperparameters that determine a structure of a machine learning network;

m represents a sample size of a mini batch Stochastic Gradient Descent (SGD) method;

X represents a $m \times k$ sample matrix;

each row of the matrix X represents a sample;

X_i represents an i^{th} row of the matrix X ;

$[j]$ represents a j^{th} element of vector X_i ;

X^L represents a first share of the sample data X that belongs to a first SCN of the plurality of secure computation nodes;

X^R represents the second share of the sample data X that belongs to a second SCN of the plurality of secure computation nodes;

Y represents a prediction value;

Y_i represents a label of the i^{th} sample of the mini batch SGD method;

E represents an error vector;

R , L , and T represent indexes of the shares of a respective SCN; and

Θ represents a parameter column vector, and the column vector Θ that is vertically dividable into Θ^L and Θ^R .

19. The system of claim 18, wherein calculating a share of A representing a column vector comprises:

encrypting original data provided by each SCN by using random numbers, and

exchanging the encrypted data between the plurality of secure computation

nodes.

20. The system of claim 19, wherein each of the random numbers is at least one of a number, a vector, or a matrix.