

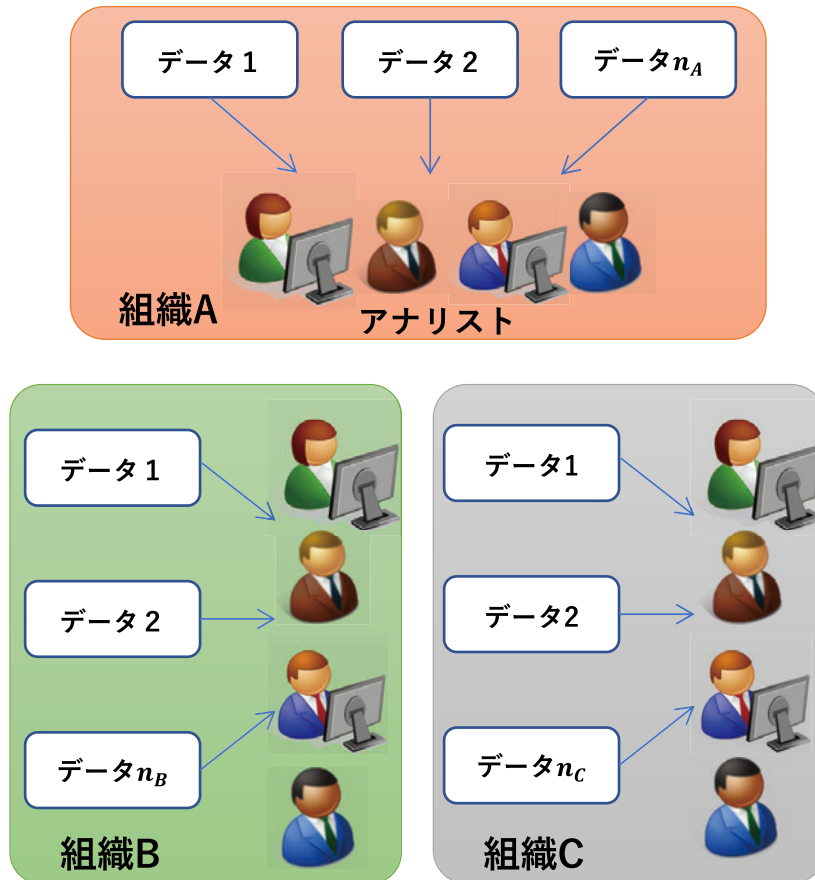
プライバシー保護データ分析の最新動向

サイバーセキュリティ研究所
セキュリティ基盤研究室
レ チュウ フォン

NICTサイバーセキュリティシンポジウム
2018年 2月14日

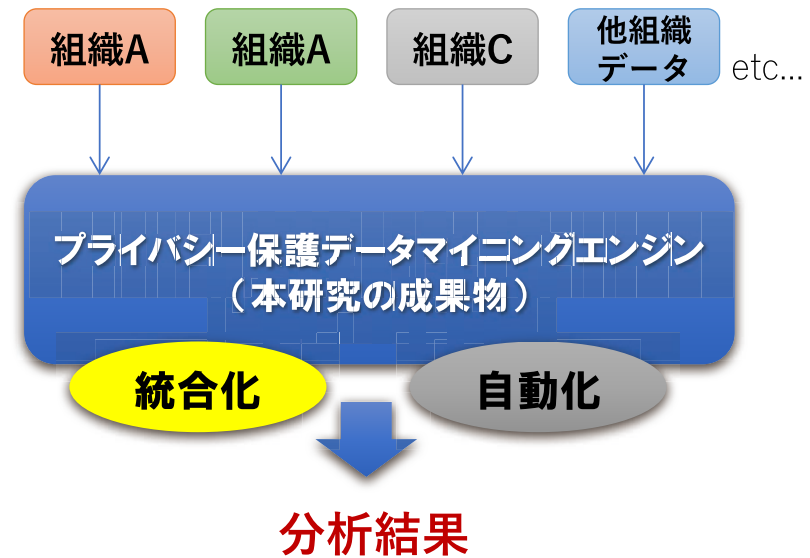
プライバシー保護データ分析

現状



個々の組織内で分析

めざす構想



- データを開示せずに、良い分析結果が得られる
- 応用：
銀行の顧客のデータ、病院の患者のデータなど

NICTでの本テーマに関する取り組み

2015年：暗号化したまま、線形回帰計算

- 100万件のデータを30分程度で処理
- 報道発表2015年1月19日。論文誌IEICE Trans. 2017

2016年：暗号化したまま、ロジスティック回帰計算

- 1億件のデータを30分以内で処理
- 報道発表2016年1月14日。論文誌IEICE Trans. 2016

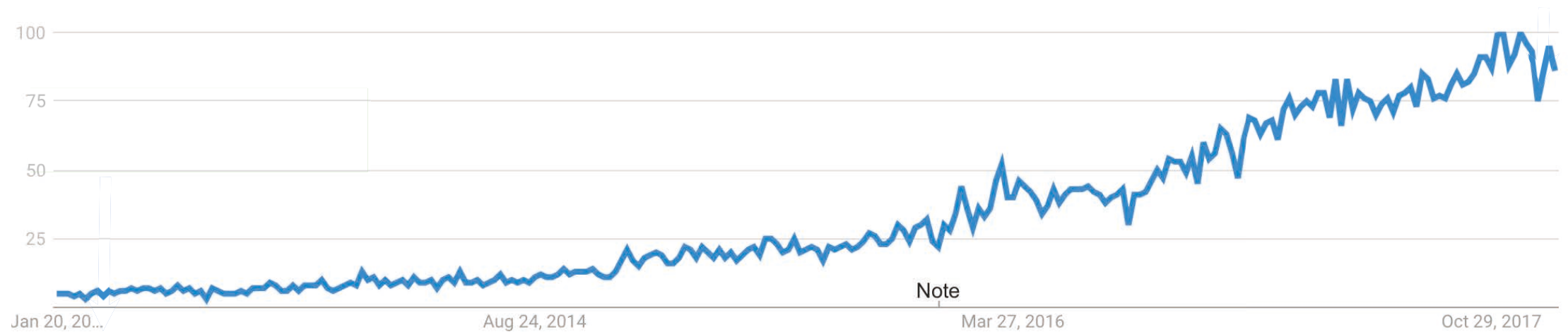
2017、2018年：暗号化したまま、ディープラーニング

- 28万件の金融データを数分で処理
- 国際会議ATIS 2017、Sec. Big Data 2017、論文誌IEEE Trans. 2018
- 引き続き取り組んでいる

今日のトピック

プライバシー保護ディープラーニング

Google検索からみた ディープラーニングへの関心度

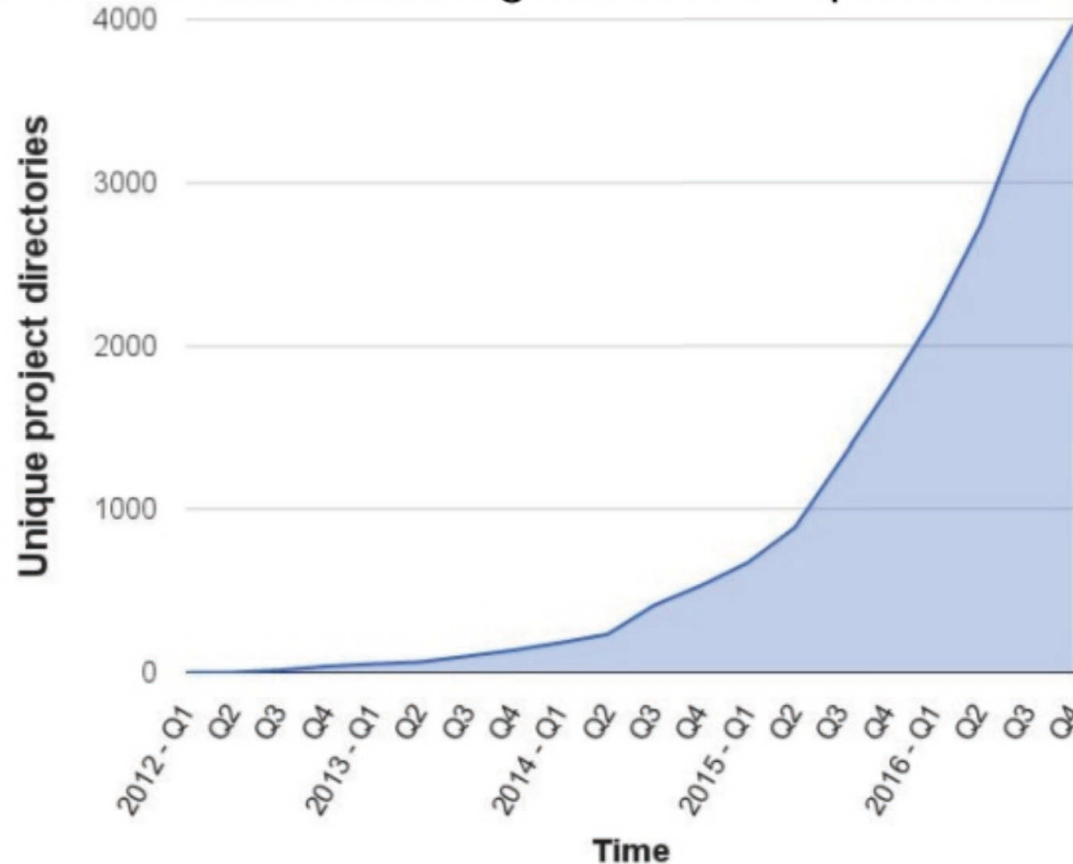


<https://trends.google.com/trends/explore?date=today%205-y&q=deep%20learning>

Google社内におけるディープラーニングを使ったプロジェクト数

Growing Use of Deep Learning at Google

of directories containing model description files



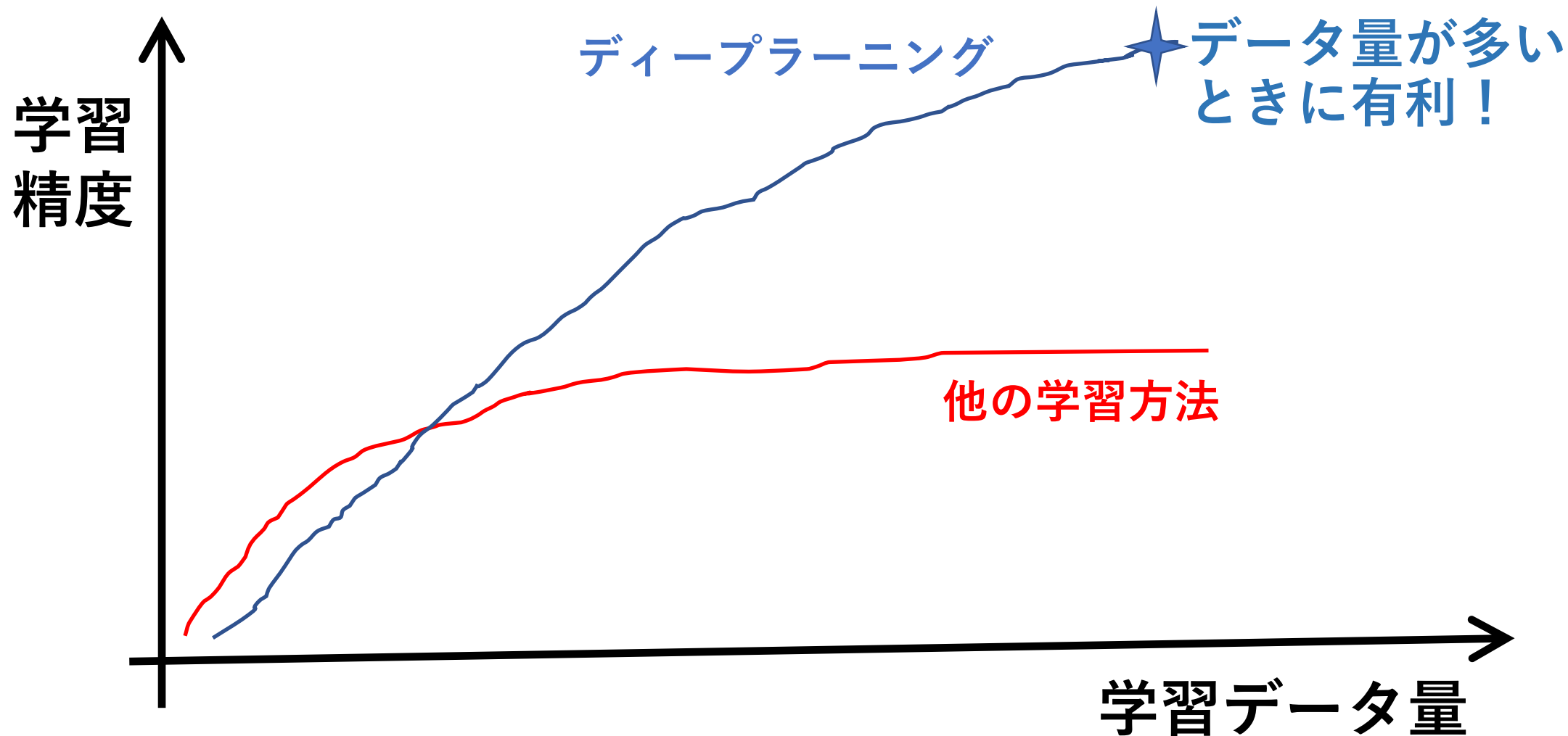
Across many products/areas:

Android
Apps
drug discovery
Gmail
Image understanding
Maps
Natural language understanding
Photos
Robotics research
Speech
Translation
YouTube
... many others ...

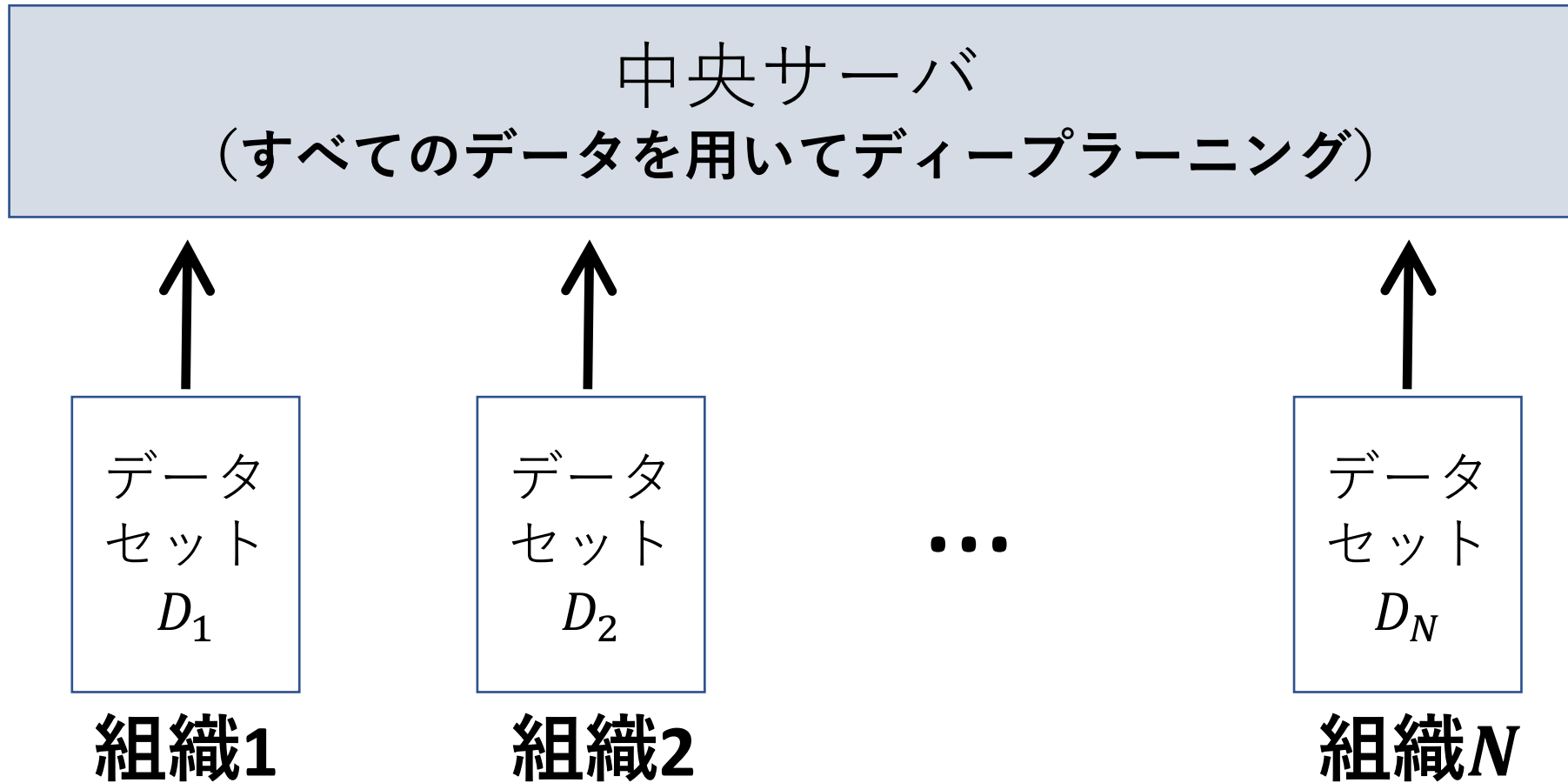
スライドソース：

<https://www.slideshare.net/AIFrontiers/jeff-dean-trends-and-developments-in-deep-learning-research>

何故ディープラーニングが使われるのか？

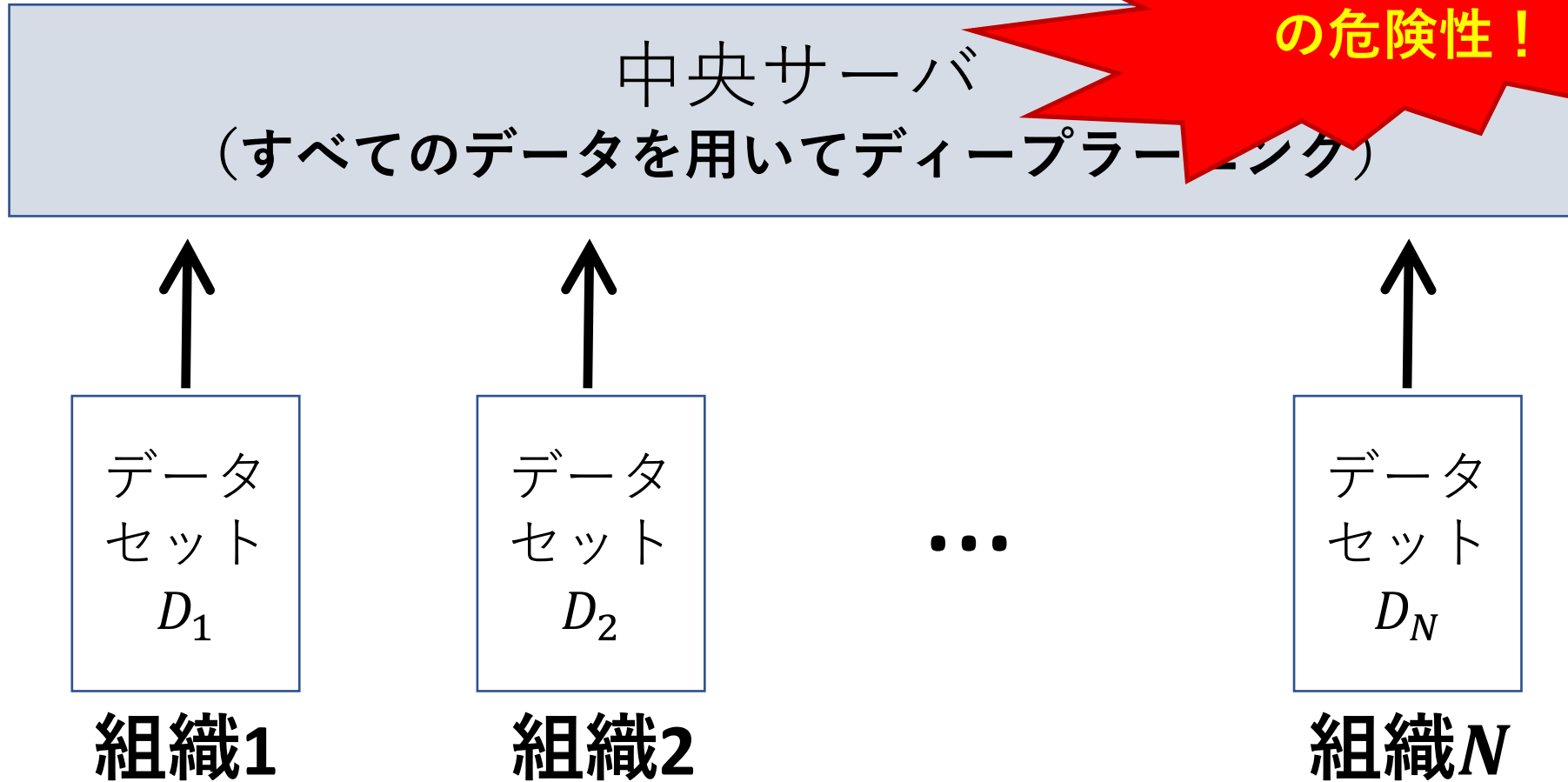


データ量を増やすための単純な方法



※組織の例：金融機関、病院

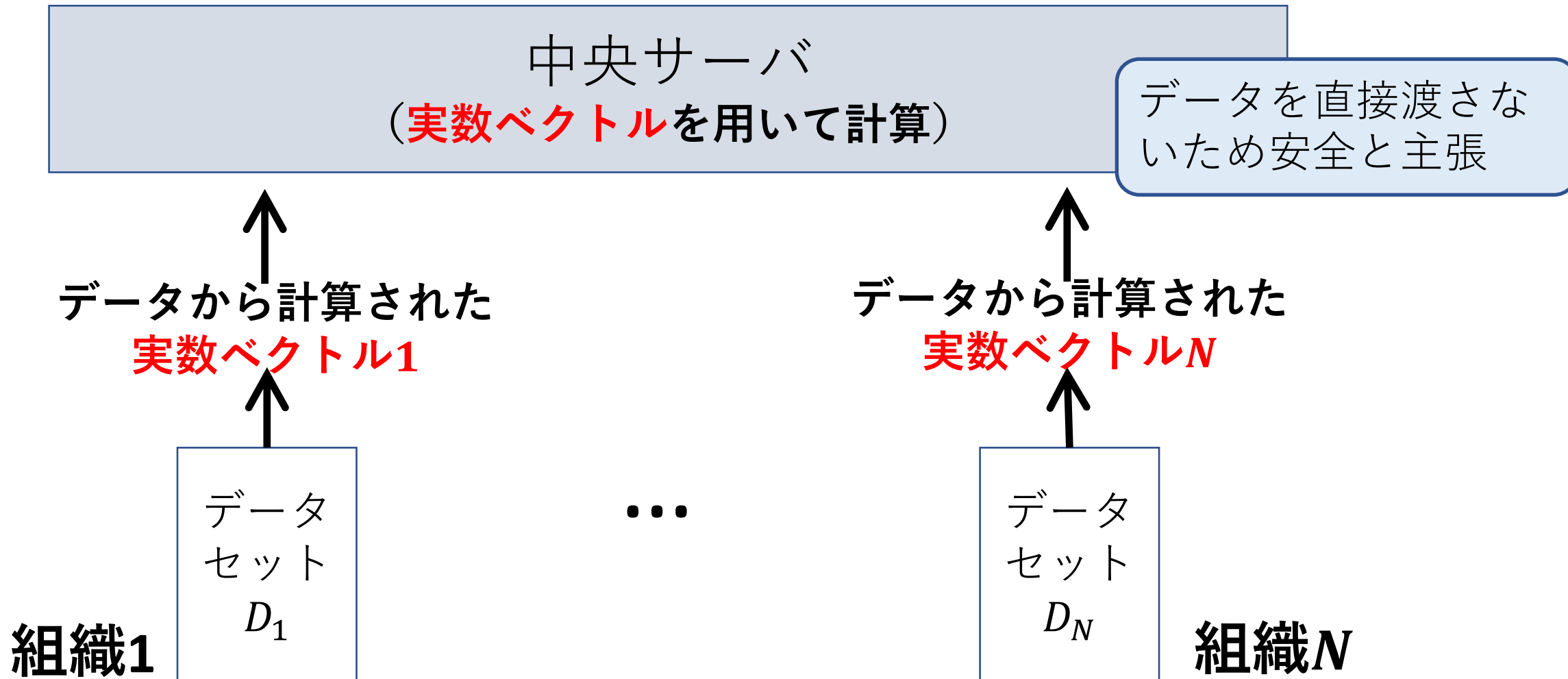
データ量を増やすための単純な方法



※組織の例：金融機関、病院

先行研究: プライバシー保護ディープラーニング

R. Shokri, S. Shmatikov: Privacy-preserving deep learning, ACM CCS 2015



NICTの研究: プライバシー保護ディープラーニング (国際会議ATIS 2017,およびSec. Big Data 2017で発表済み)



1. 先行研究の問題点を分析

Shokri-Shmatikov2015の方式では、中央サーバに**データが漏洩する**

2. 解決方法の提案

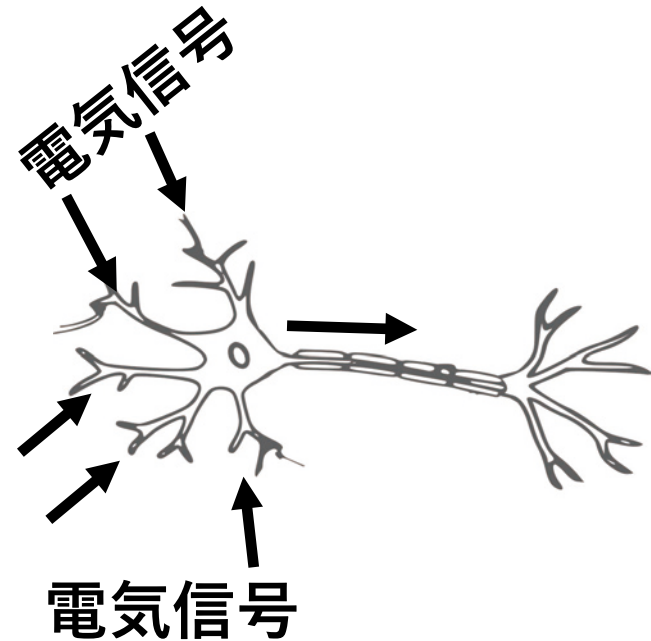
新しいプライバシー保護ディープラーニングシステム：

- a) **加法準同型暗号**を使うシステム (Shokri-Shmatikov2015研究をベースに修正)
- b) **共通鍵暗号**を使うシステム (独自の提案)

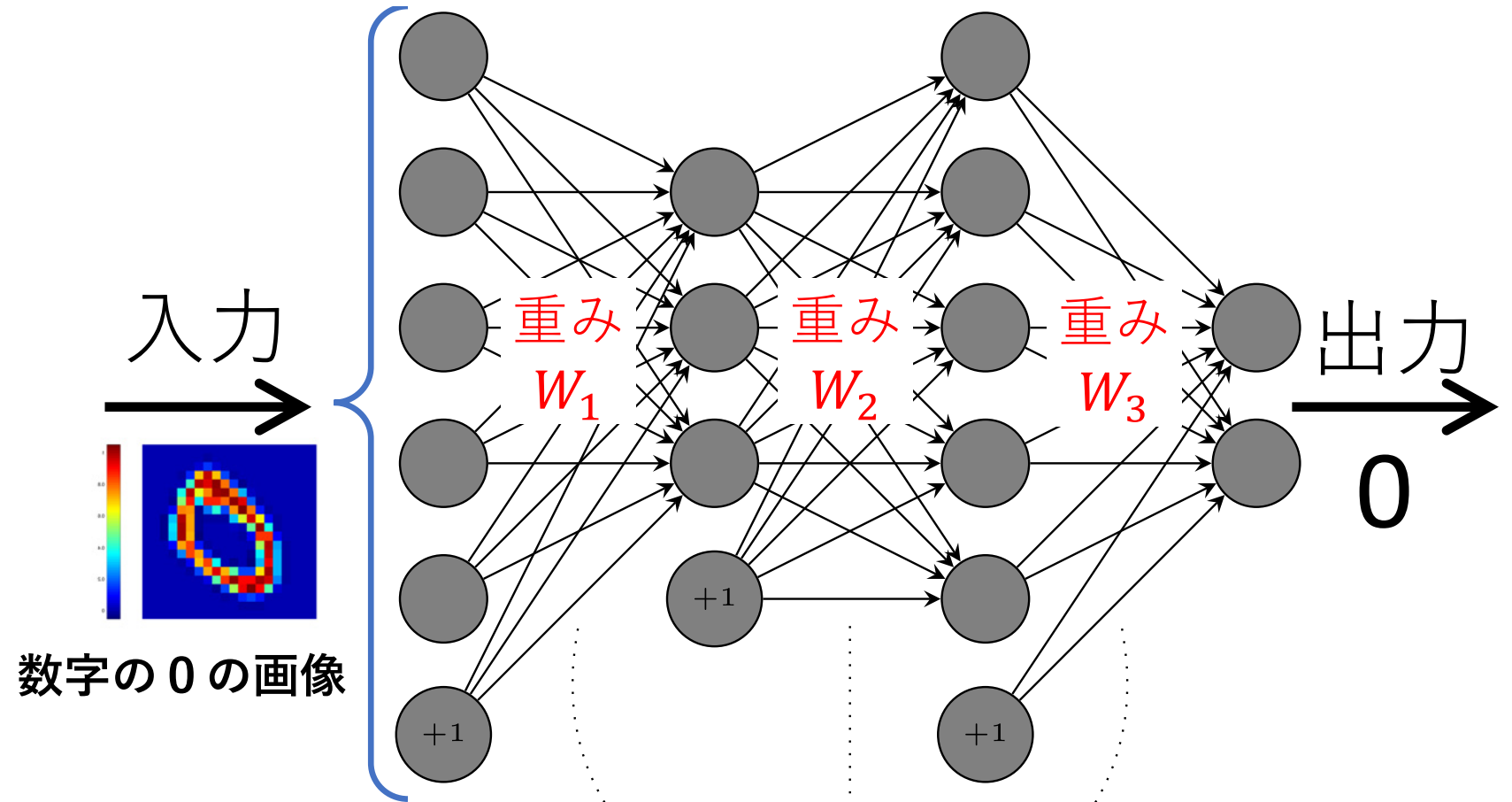
- a) L. T. Phong et al.: Privacy-Preserving Deep Learning: Revisited and Enhanced. ATIS 2017, IEEE Trans. Information Forensics & Security 2018
- b) L. T. Phong: Privacy-Preserving Stochastic Gradient Descent with Multiple Distributed Trainers. Sec. Big Data@NSS 2017: 510-518

準備 1/2 : ニューラルネットワーク

生物の神経細胞 (ニューロン)



人工知能 (AI) のニューラルネットワーク



ニューラルネットの学習： 大量のデータ (= 入出力のペア) に対して、できる限りそれを再現するように **重みベクトル W** を調整

準備 2 / 2 : 確率的勾配降下法(SGD)

関数 $J(W, data)$: 現在の重み W を用いた
ニューラルネットによるデータ再現性の指標

重みベクトル W をランダムに初期する

REPEAT

FOR $data$ IN Dataset:

$G \leftarrow \frac{\delta(J(W, data))}{\delta W}$ (勾配 G を計算)

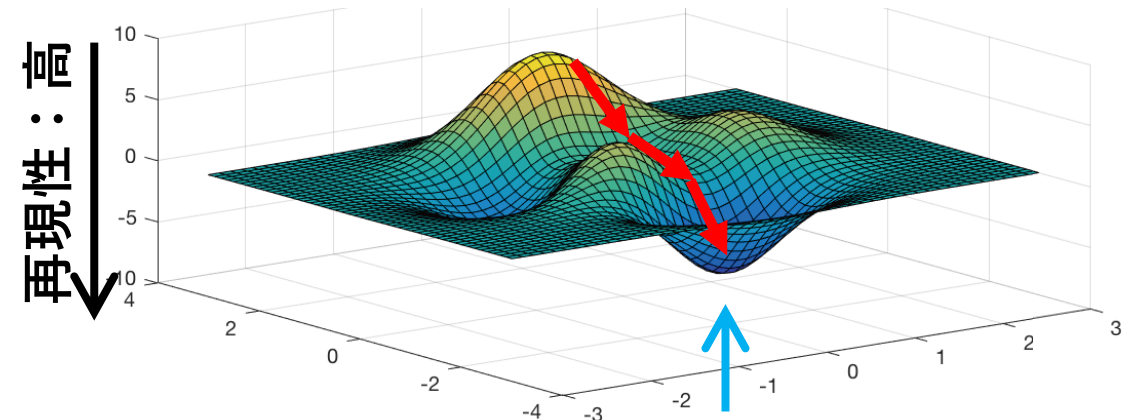
$W \leftarrow W - \alpha \cdot G$ (重みベクトルの更新)

ENDFOR

UNTIL (good accuracy)

確率的勾配降下法(SGD)

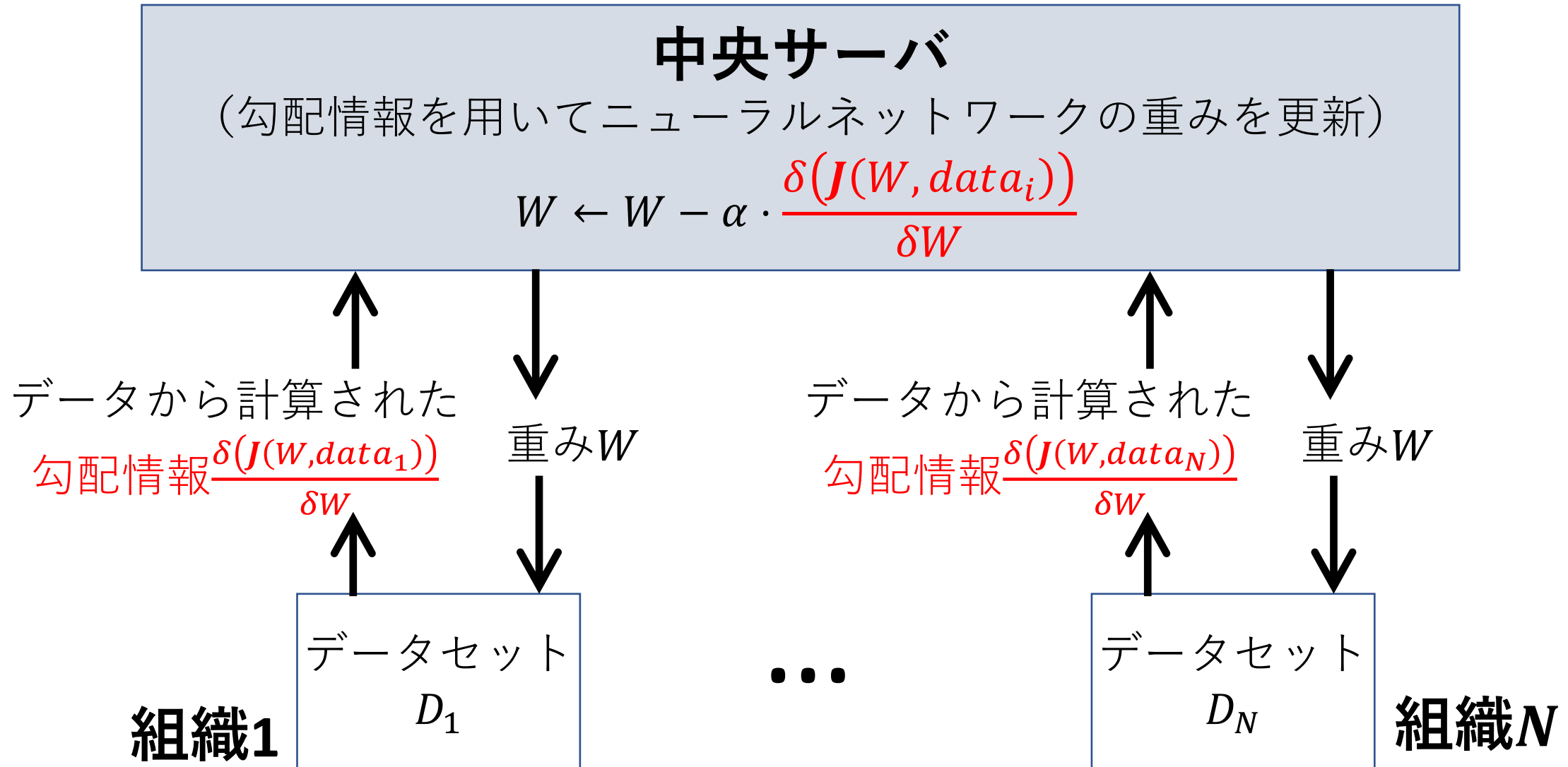
関数 $J(W, data)$



関数の勾配 G の方向に進む

プライバシー保護ディープラーニング（先行研究、詳細）

Shokri-Shmatikov, "Privacy-preserving deep learning", ACM CCS 2015



NICTの研究: プライバシー保護ディープラーニング (国際会議ATIS 2017,およびSec. Big Data 2017で発表済み)



1. 先行研究の問題点を分析

Shokri-Shmatikov2015の方式では、中央サーバに**データが漏洩する**

2. 解決方法の提案

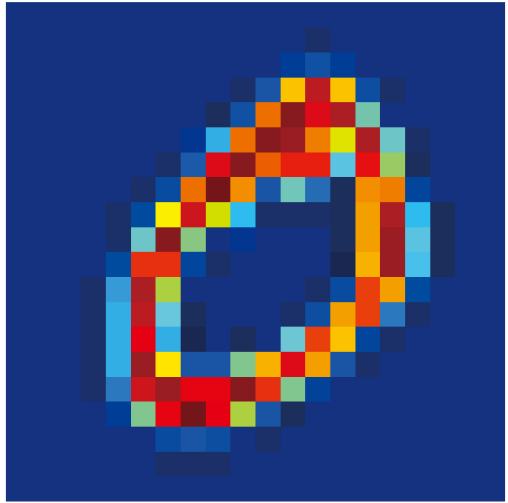
新しいプライバシー保護ディープラーニングシステム：

- a) 加法準同型暗号を使うシステム (Shokri-Shmatikov2015研究をベースに修正)
- b) 共通鍵暗号を使うシステム (独自の提案)

- a) L. T. Phong et al.: Privacy-Preserving Deep Learning: Revisited and Enhanced. ATIS 2017, IEEE Trans. Information Forensics & Security 2018
- b) L. T. Phong: Privacy-Preserving Stochastic Gradient Descent with Multiple Distributed Trainers. Sec. Big Data@NSS 2017: 510-518

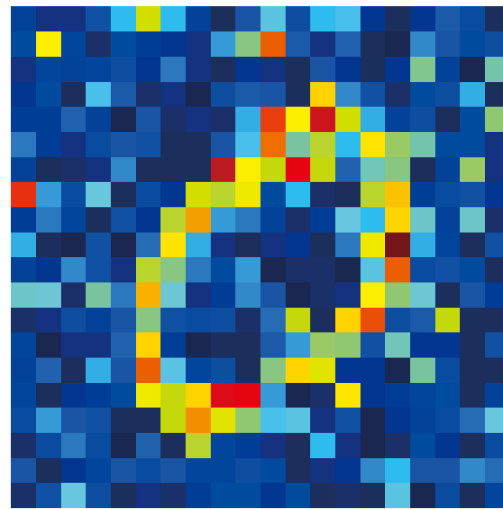
先行研究の問題点：勾配情報からのデータ漏洩

← 20ピクセル →



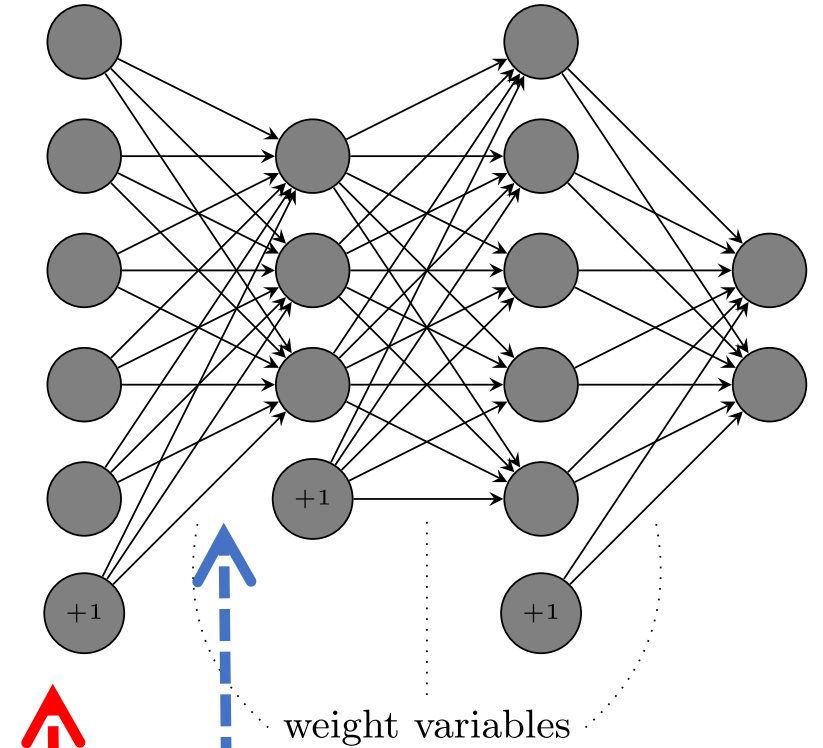
元の画像
(データ)

← 20ピクセル →



勾配情報から
計算された画像

勾配情報から、データ漏洩の危険性



$$\xi_i \cdot \mathbf{x}_k = \frac{\delta J(W, x, y)}{\delta W_{ik}^{(1)}} + \text{誤差}$$

(1 ≤ k ≤ 20×20)

NICTの研究: プライバシー保護ディープラーニング (国際会議ATIS 2017, およびSec. Big Data 2017で発表済み)



1. 先行研究の問題点を分析

Shokri-Shmatikov2015の方式では、中央サーバにデータが漏洩する

2. 解決方法の提案

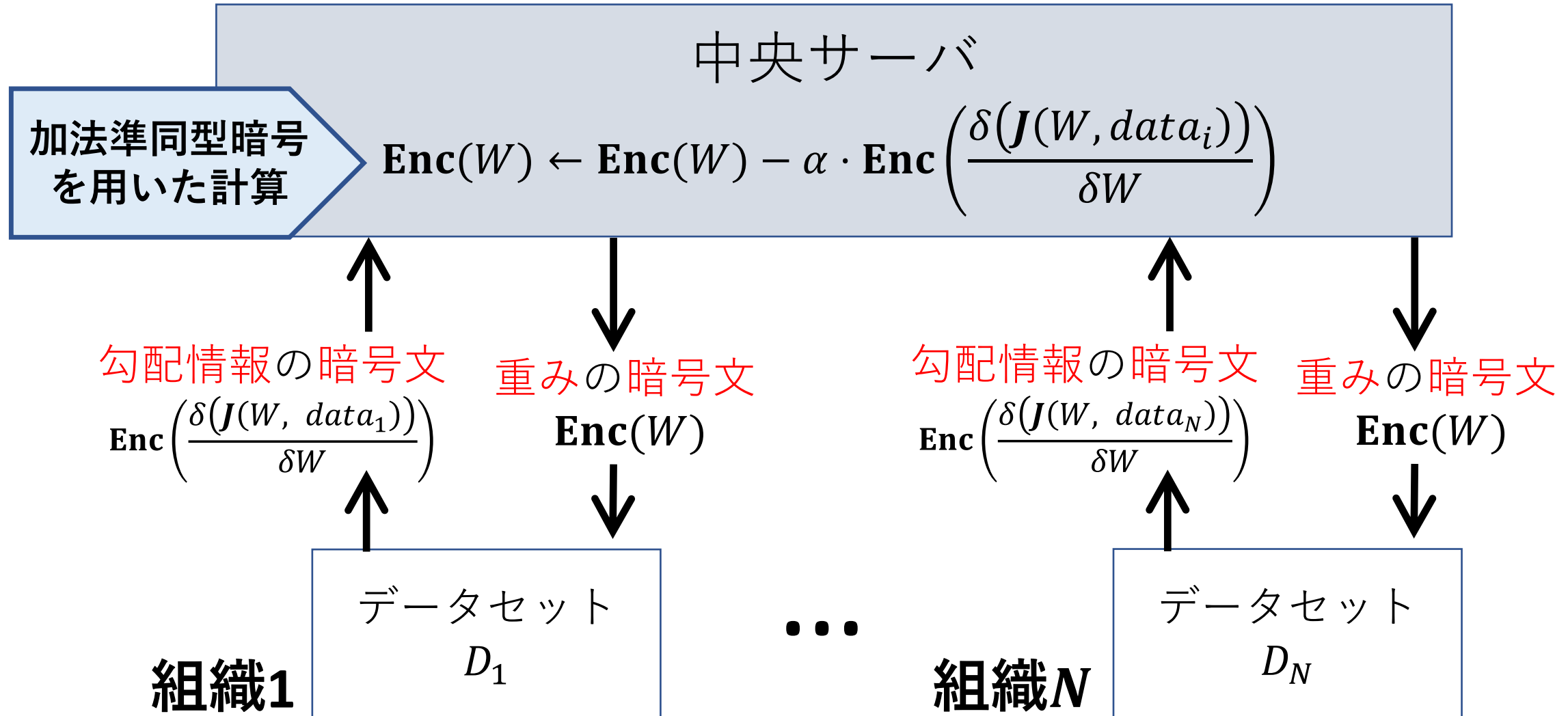
新しいプライバシー保護ディープラーニングシステム：

- a) **加法準同型暗号**を使うシステム (Shokri-Shmatikov2015研究をベースに修正)
- b) **共通鍵暗号**を使うシステム (独自の提案)

- a) L. T. Phong et al.: Privacy-Preserving Deep Learning: Revisited and Enhanced. ATIS 2017, IEEE Trans. Information Forensics & Security 2018
- b) L. T. Phong: Privacy-Preserving Stochastic Gradient Descent with Multiple Distributed Trainers. Sec. Big Data@NSS 2017: 510-518

研究成果1：加法準同型暗号を用いたシステム

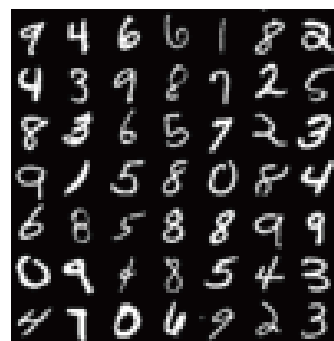
- 勾配情報を暗号化すれば、データ漏洩は防げる
- 中央サーバにおける計算は加算のみなので、暗号化したまま加算ができれば良い



通信量評価

暗号方式	通信料の増加（倍）
SPHERE方式（NICT開発）	2.46 (MNIST); 2.43 (SVHN); 2.41 (speech dataset)
Paillier暗号	2.93 (all datasets)

MNIST: Modified National Institute of Standards and Technology. **SVHN**: Street View House Number



通信量の例: MNIST dataset

- 暗号化なし： 0.56 MB (1回の重み更新)
- 我々の研究成果1： $0.56 \times \mathbf{2.46} = 1.38$ MB (1回の重み更新)

計算量の評価

MNISTデータセット使用

ニューラル ネットワーク	学習変数の数	学習の精度	暗号化なし	準同型暗号 を利用した提案システム 1
多層 パーセプトロン	109,386	≈ 97%	2 分程度	2.3 時間以上

時間がかかった理由：

- 10万個の実数を暗号化（復号化）する時間：190 ミリ秒（一回）
- 学習に使う暗号化（復号化）の回数：20,000 回
- 時間の合計：190 ミリ秒 $\times 20,000 \times 2 = 2.12$ 時間

改良のアプローチ：

- **アプローチ 1**：中央サーバでの並列化、暗号パラメータの精密、実装の最適化
- **アプローチ 2**：プロトコル設計を見直す→ **次のスライドで紹介**

準備 2 / 2 : 確率的勾配降下法(SGD)

関数 $J(W, data)$: 現在の重み W を用いた
ニューラルネットによるデータ再現性の指標

重みベクトル W をランダムに初期する

REPEAT

FOR $data$ IN Dataset:

$G \leftarrow \frac{\delta(J(W, data))}{\delta W}$ (勾配 G を計算)

$W \leftarrow W - \alpha \cdot G$ (重みベクトルの更新)

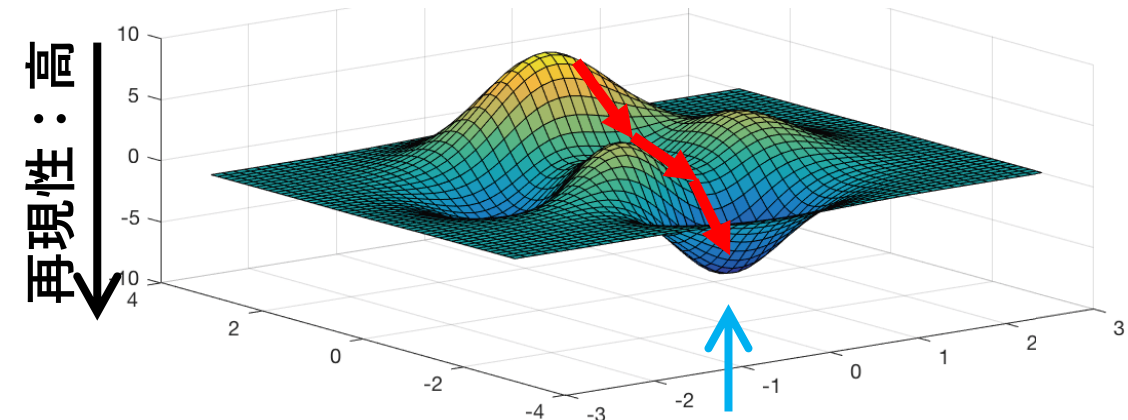
ENDFOR

RETURN W

UNTIL (good accuracy)

確率的勾配降下法(SGD)

関数 $J(W, data)$

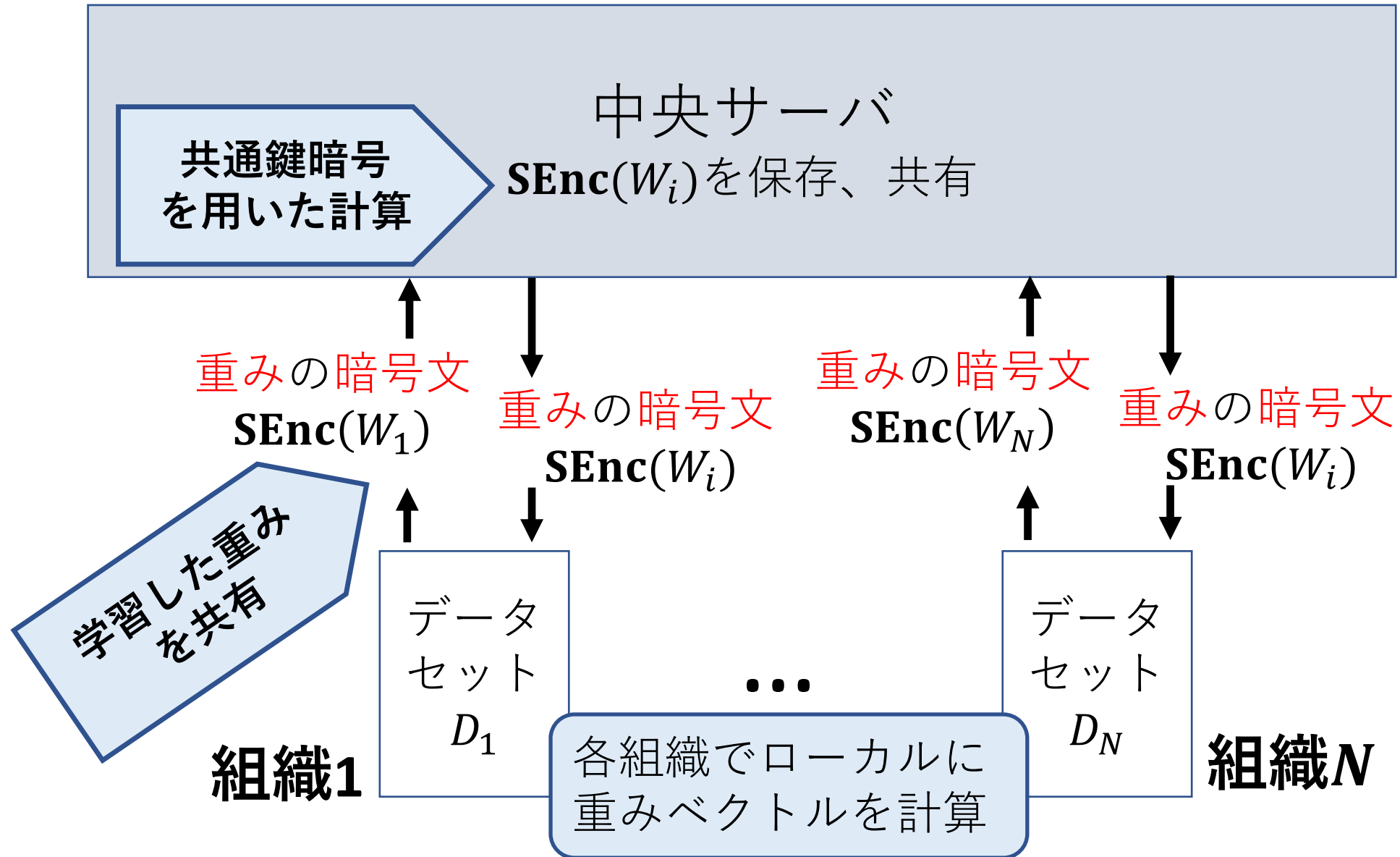


関数の勾配 G の方向に進む

勾配 G を暗号化すると暗号化の回数が多い → 時間がかかる！

勾配の代わりに、重み W を暗号化すれば、暗号化の回数が減る

研究成果2：共通鍵暗号を使うシステム



計算量の評価（MNISTデータセット使用）

ニューラルネットワーク	学習変数の数	学習の精度	暗号化なし	提案システム 学習の処理時間
多層パーセプトロン	109,386	≈ 97%	2 分程度	2.3 時間以上 (研究成果1を使用)
畳み込み	1,199,882	≈ 99%	3分 (一台のサーバ) 10分 (分散協調学習)	0.5 時間 (研究成果2を使用)

※学習の後に、予測の時間：

ー各組織はローカルで行えるので、高速にできる。

ー例：1万件のMNIST画像の予測は、3秒程度。

研究成果の比較

方式	中央サーバ に対する安全性	学習結果 の正確さ	暗号技術
従来研究 (SS2015)	×	非同期SGD と同等	なし
NICT成果1	○	非同期SGD と同等	加法準同型暗号 (勾配情報を暗号化)
NICT成果2	○	同期SGD と同等	共通鍵暗号 (重みを暗号化)

※SGD: 確率的勾配降下法

オープンデータを用いたプロトタイプの検証

対象：研究用の公開データ

- ヨーロッパのクレジットカード会員の2日間の取引データ
- 284,807件の取引レコード（内0.2%程度が不正利用に関わる）

目的：取引レコードから、不正利用を検出

<https://www.kaggle.com/dalpozz/creditcardfraud>

Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson and Gianluca Bontempi. Calibrating Probability with Undersampling for Unbalanced Classification. In Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2015

オープンデータを用いたプロトタイプの検証

複数組織からのデータ提供をシミュレーションするため、
公開データを計21個のグループに分割

- 学習用20セット + テスト用1セット

	不正利用	正当な利用
学習用データセット 1,...,20	13 ~ 27	≈ 11379
テストセット	98	56863

プロトタイプ

Fraud Detector Collaborative Machine Training		
F-score 0.00000	Round 1	
Accuracy 0.99828		
Data Owner 1	SERVER	Data Owner 11
TP=0, FN=98, FP=0, TN=56863	M4_NK5	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	4_NK5k	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	_NK5kh	
Data Owner 2	NK5khR	Data Owner 12
TP=0, FN=98, FP=0, TN=56863	K5khRi	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	5khRi_	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	khRi_l	
Data Owner 3	hRi_lI	Data Owner 13
TP=0, FN=98, FP=0, TN=56863	Ri_lIq	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	i_lIqJ	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	_lIqJr	
Data Owner 4	lIqJrZ	Data Owner 14
TP=0, FN=98, FP=0, TN=56863	IqJrZ6	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	qJrZ6f	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	JrZ6fv	
Data Owner 5	rZ6fvN	Data Owner 15
TP=0, FN=98, FP=0, TN=56863	Z6fvNv	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	6fvNvR	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	fvNvRy	
Data Owner 6	vNvRyy	Data Owner 16
TP=0, FN=98, FP=0, TN=56863	NvRyyg	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	vRyygn	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	RyygnF	
Data Owner 7	yygnFV	Data Owner 17
TP=0, FN=98, FP=0, TN=56863	ygnFVq	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	gnFVq3	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	nFVq3K	
Data Owner 8	FVq3K8	Data Owner 18
TP=0, FN=98, FP=0, TN=56863	Vq3K8m	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	q3K8m8	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	3K8m8X	
Data Owner 9	K8m8X1	Data Owner 19
TP=0, FN=98, FP=0, TN=56863	m8X1M	TP=0, FN=98, FP=0, TN=56863
xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	m8X1MH	xxxxxxxxxxxxxxxxxxxxxxxxxxxx 0.0%
	8X1MH	

ラウンド1



中央サーバ

データ所有者
(計20組織)

データ所有者
(計20組織)

学習ラウンド1終了時には、テストセット内の98件の不正利用が
検出できていない

プロトタイプ

Fraud Detector Collaborative Machine Training		
F-score 0.85567	Round 3	
Accuracy 0.99951		
Data Owner 1	SERVER	Data Owner 11
TP=0, FN=98, FP=0, TN=56863	UhDeg0	TP=75, FN=23, FP=10, TN=56853
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	hDeg0R	xxxxxxxxxxxxxxx0000000000 51.0%
Data Owner 2	Deg0Rk	
TP=0, FN=98, FP=0, TN=56863	eg0Rka	Data Owner 12
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	g0RkaV	TP=77, FN=21, FP=10, TN=56853
Data Owner 3	0RkaVV	xxxxxxxxxxxxxxx0000000000 66.3%
TP=0, FN=98, FP=0, TN=56863	RkaVV2	
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	kaVV2I	Data Owner 13
Data Owner 4	aVV2Iy	TP=73, FN=25, FP=10, TN=56853
TP=0, FN=98, FP=0, TN=56863	VV2Iyb	xxxxxxxxxxxxxxx0000000000 68.4%
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	V2Iyb1	
Data Owner 5	2Iyb10	Data Owner 14
TP=0, FN=98, FP=0, TN=56863	Iyb10A	TP=75, FN=23, FP=10, TN=56853
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	yb10Ak	xxxxxxxxxxxxxxx0000000000 64.3%
Data Owner 6	b10AkB	
TP=0, FN=98, FP=0, TN=56863	10AkBA	Data Owner 15
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	0AkBAh	TP=83, FN=15, FP=13, TN=56850
Data Owner 7	AKBAhj	xxxxxxxxxxxxxxx000000000000 66.5%
TP=0, FN=98, FP=0, TN=56863	kBAhjb	
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	BAhjbz	Data Owner 16
Data Owner 8	Ahjbza	TP=81, FN=17, FP=11, TN=56852
TP=0, FN=98, FP=0, TN=56863	hjbza-	xxxxxxxxxxxxxxx00000000000000 71.4%
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	jbza-3	
Data Owner 9	bza-38	Data Owner 17
TP=0, FN=98, FP=0, TN=56863	za-382	TP=81, FN=17, FP=12, TN=56851
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	a-382j	xxxxxxxxxxxxxxx00000000000000 71.4%
Data Owner 10	-382j9	
TP=0, FN=98, FP=0, TN=56863	382j9d	Data Owner 18
xxxxxxxxxxxxxxxxxxxxxxxxx 0.0%	82j9dA	TP=83, FN=15, FP=13, TN=56850
	2j9dAI	xxxxxxxxxxxxxxx000000000000000 70.4%
	j9dAID	
	9dAIDo	Data Owner 19
	dAIDoH	TP=83, FN=15, FP=13, TN=56850
	AIDoH1	xxxxxxxxxxxxxxx000000000000000 71.4%
	IDoH1B	
	DoH1BC	Data Owner 20
	oH1BCC	TP=0, FN=98, FP=0, TN=56863
	H1BCCc	

ラウンド3



中央サーバ

データ所有者
(計20組織)

データ所有者
(計20組織)

学習が進むにつれて、不正利用の検出率が上がっていく

プロトタイプ

[phong@vt001 credit_card]\$ Fraud Detector Collaborative Machine Training
F-score max 0.86154
Accuracy max 0.99953

ラウンド6



データ
所有者
(計20組織)

Round 6		
SERVER		
Data Owner 1	4UuD_V	Data Owner 11
TP=85, FN=13, FP=18, TN=56845	UuD_VE	TP=86, FN=12, FP=19, TN=56844
xxxxxxx0000000000000000 68.4%	uD_VEL	xxxxxxx0000000000000000 69.4%
	D_VELR	
Data Owner 2	_VELRV	Data Owner 12
TP=85, FN=13, FP=18, TN=56845	VELRVW	TP=85, FN=13, FP=18, TN=56845
xxxxxxx0000000000000000 68.4%	ELRVW5	xxxxxxx0000000000000000 68.4%
	LRVW5K	
Data Owner 3	RVW5Kj	Data Owner 13
TP=85, FN=13, FP=18, TN=56845	VW5Kja	TP=84, FN=14, FP=17, TN=56846
xxxxxxx0000000000000000 68.4%	W5Kjaq	xxxxxxx0000000000000000 68.4%
	5KjaqV	
Data Owner 4	KjaqVQ	Data Owner 14
TP=85, FN=13, FP=18, TN=56845	jaqVQA	TP=84, FN=14, FP=17, TN=56846
xxxxxxx0000000000000000 68.4%	aqVQA1	xxxxxxx0000000000000000 68.4%
	qVQA1S	
Data Owner 5	VQA1SB	Data Owner 15
TP=85, FN=13, FP=18, TN=56845	QA1SBh	TP=85, FN=13, FP=18, TN=56845
xxxxxxx0000000000000000 68.4%	A1SBhe	xxxxxxx0000000000000000 68.4%
	1SBheZ	
Data Owner 6	SBheZt	Data Owner 16
TP=85, FN=13, FP=18, TN=56845	BheZtI	TP=84, FN=14, FP=17, TN=56846
xxxxxxx0000000000000000 68.4%	heZtIq	xxxxxxx0000000000000000 68.4%
	eZtIqI	
Data Owner 7	ZtIqIy	Data Owner 17
TP=84, FN=14, FP=17, TN=56846	tIqIyU	TP=83, FN=15, FP=15, TN=56848
xxxxxxx0000000000000000 68.4%	IqIyU2	xxxxxxx0000000000000000 68.4%
	qIyU2g	
Data Owner 8	IyU2gg	Data Owner 18
TP=85, FN=13, FP=18, TN=56845	yU2ggE	TP=84, FN=14, FP=17, TN=56846
xxxxxxx0000000000000000 68.4%	U2ggEa	xxxxxxx0000000000000000 69.4%
	2ggEap	
Data Owner 9	ggEapP	Data Owner 19
TP=85, FN=13, FP=18, TN=56845	gEapPC	TP=85, FN=13, FP=18, TN=56845

中央サーバ

データ
所有者
(計20組織)

学習アルゴリズムの収束後、98件中85件の不正利用が検出可能になっている

- I. プライバシー保護ディープラーニングシステムを提案
 - a) 加法準同型暗号を使うシステム
 - b) 共通鍵暗号を使うシステム
- II. プロトタイプを実装し、クレジットカード取引の不正利用検出への応用が可能であることを実証

ご清聴、ありがとうございました