Privacy-Preserving Deep Learning

Reza Shokri and Vitaly Shmatikov

(UT Austin and Cornell Tech)

Machine Intelligence LANDSCAPE

CORE TECHNOLOGIES

ARTIFICIAL INTELLIGENCE
IBM WATSON · MetaMind · Numenta · ai-one · Cycorp · Research · nara · Reactor · SI

DEEP LEARNING
vicarious · Vision Factory · facebook · Baidu IDL · Google · ersatz · SKYMIND · SignalSense

MACHINE LEARNING
rapidminer · context · Oxdata H2O · DATARPM · Alpine · Sonian

NLP PLATFORMS
cortical.io · idibon · LUMINOSO · wit.ai · Maluuba

PREDICTIVE APIS
AlchemyAPI · MINDOPS · Google · bigml · indico · ALGORITHMIA · Expect Labs · PredictionIO

IMAGE RECOGNITION
clarifai · MADBITS · DNNresearch · DEXTRO · VISENZE · lookflow

SPEECH RECOGNITION
GRIDSPACE · popup archive · NUANCE

RETHINKING ENTERPRISE

SALES
Preact · AVISO · RelateIQ · NGDATA · FRAMED · infer · ACTIVITY · causata

SECURITY / AUTHENTICATION
CROSSMATCH · enjur · BITSIGHT · CYLANCE · bionym

FRAUD DETECTION
sift science · SOCURE · ThreatMetrix · feedzai · Brighterion · VERAFIN

HR / RECRUITING
TalentBin · entelo · predikt · Connectifier · gild · hi

MARKETING
brightfunnel · bloom · Commun.it

PERSONAL ASSISTANT

INTELLIGENCE TOOLS
ADATAD · Palantir · Quid · Digital Reasoning · FirstRain

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World

ADTECH
METAMARKETS · dstillery · rocketfuel · YieldMo · ADBRAIN

MEDICAL
Parvial · transcriptic · ZEPHYR · Genescient · bina · TUTE

OIL AND GAS
kaggle · AYASDI · TACHYUS · biota · Futura

FINANCE
Affirm · LendUp · LendingClub · Kabbage

PHILANTHROPIES
DataKind · thorn · DATA GUILD

AUTOMOTIVE
Google · Continental · CRUISE

DIAGNOSTICS
enlitic · 3SCAN · lumiata

RETAIL
BAY SENSORS · PRISM SKYLABS · celect · euclid

RETHINKING HUMANS / HCI

AUGMENTED REALITY
APX · blippar · META · layar

GESTURAL COMPUTING
THALMICLABS · omek · LEAP · 3Gear · eyeSight · GestureTek · nod

ROBOTICS
intel · Rethink Robotics · jibo · anki

EMOTIONAL RECOGNITION
affectiva · BEYOND VERBAL · EMOTIENT · cogito

SUPPORTING TECHNOLOGIES

HARDWARE
NVIDIA · XILINX · QUALCOMM · NERVANA · rigetti

DATA PREP
TRIFACTA · Paxata · tamr · Alation

DATA COLLECTION
diffbot · kimono · CrowdFlower · Connotate · WorkFusion · import.io

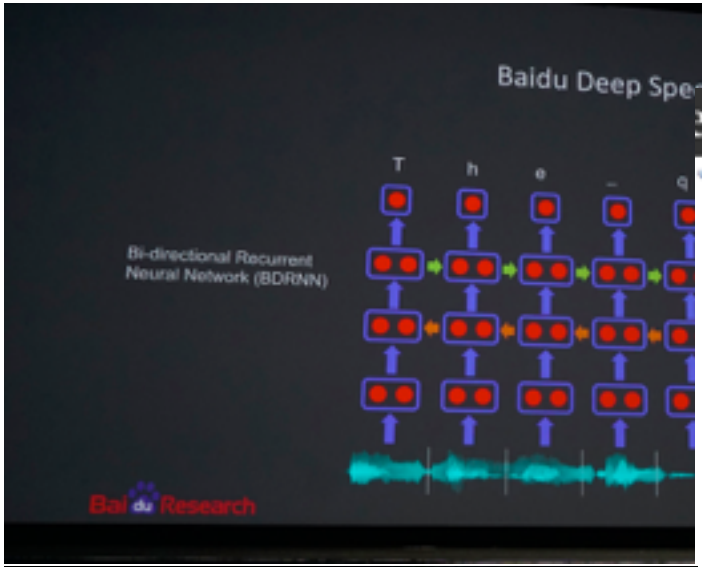
rbes ▾ New Posts +1 Most Popular Lists Video Subscribe to Forbes

Sign up

TECH 12/18/2014 @ 9:00AM | 48,377 views

Baidu Announces Breakthrough In Speech Recognition, Claiming To Top Google And Apple

Robert Hof
Contributor



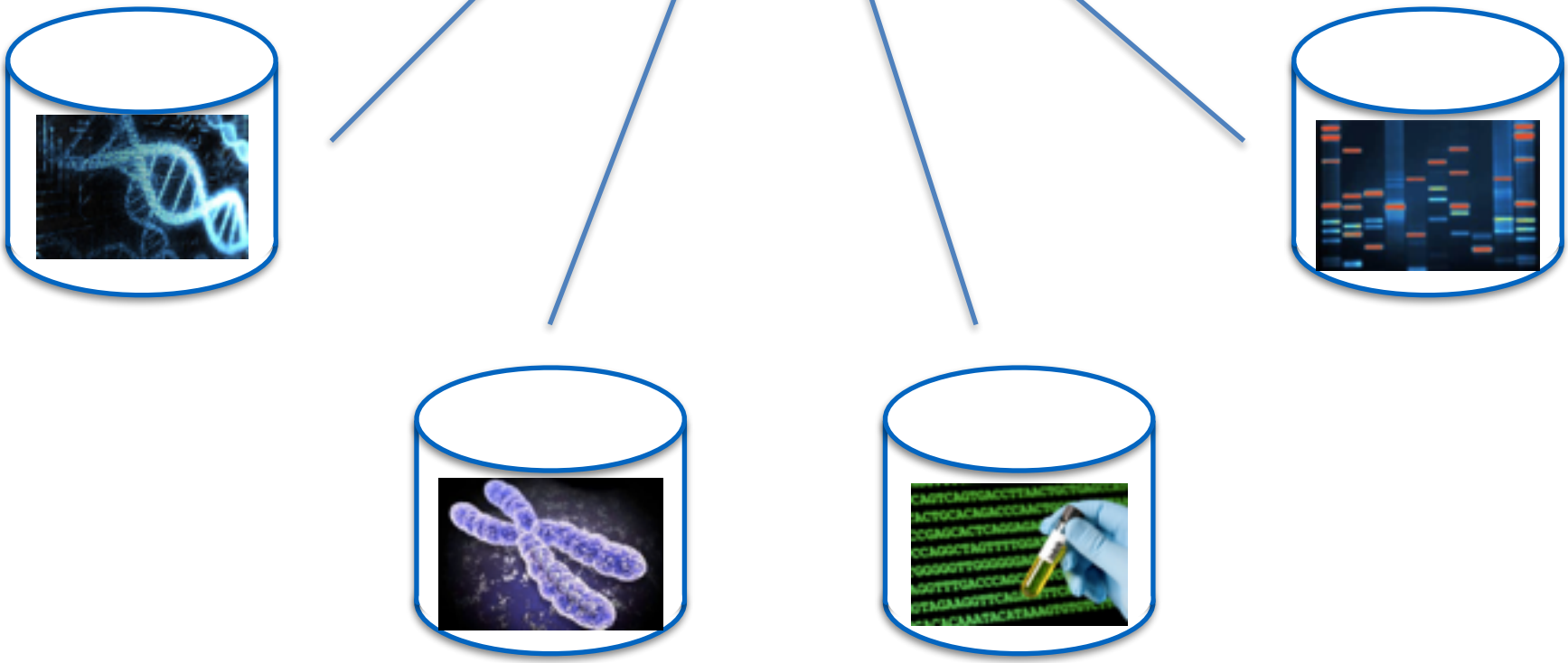
TECH 2/13/2015 @ 1:00PM | 4,996 views

Microsoft's Deep Learning Project Outperforms Humans In Image Recognition

Michael Thomsen
Contributor

Images used to test Microsoft's deep learning AI. Image via Microsoft.

Deep Learning Today



Privacy Concerns

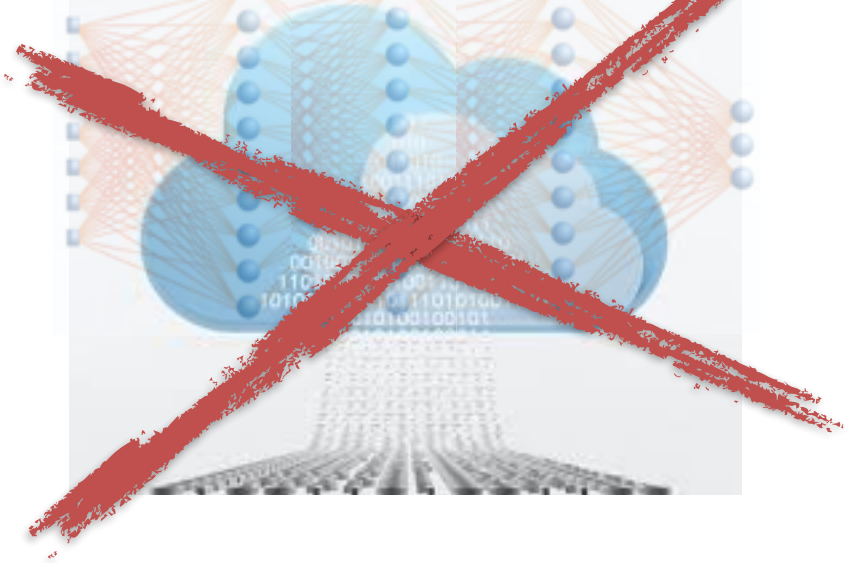
- Training data is very sensitive
  - speech, photo images, written documents, might contain sensitive private information
- Users have no control over the learning objective
- Using trained networks requires users to share their private data with service providers

Possible Consequences

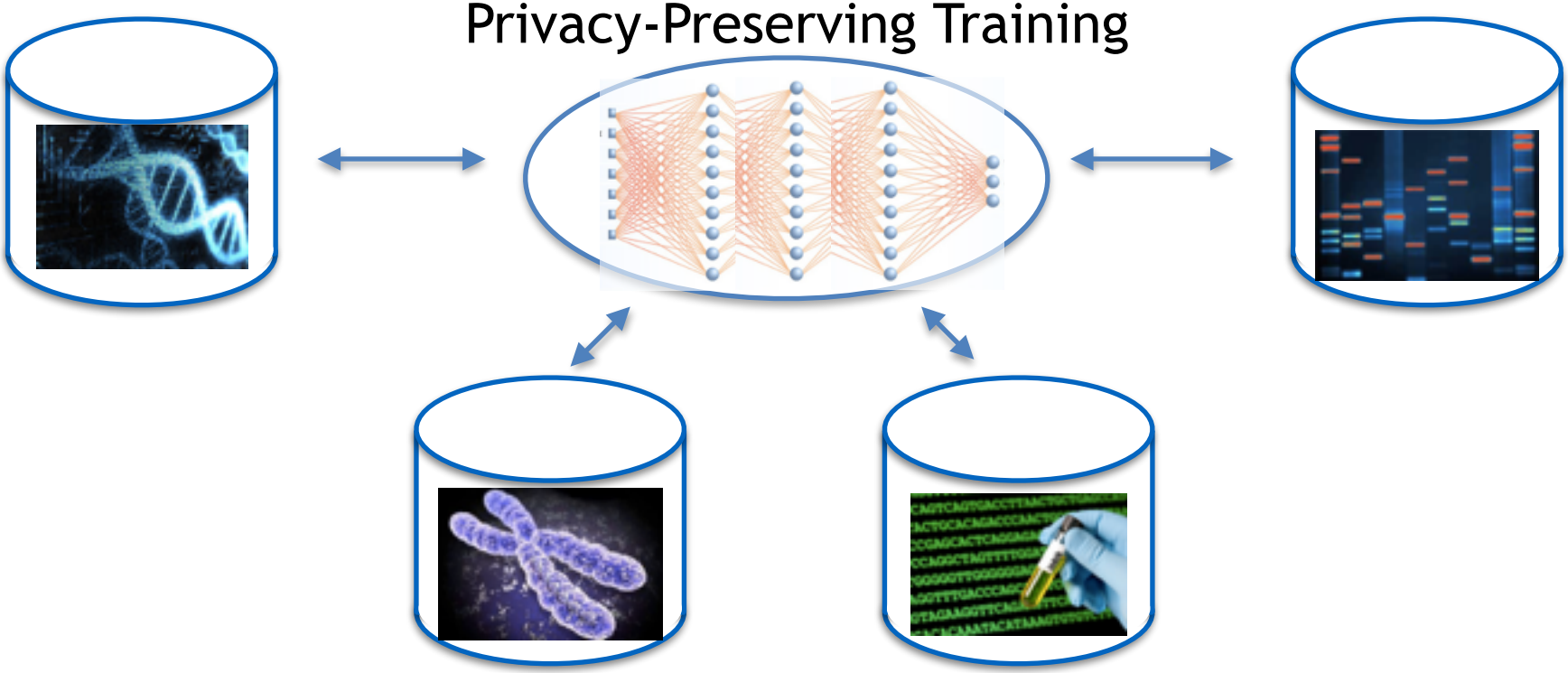
- Users' data might be used in wrong context
  - Compromises and data breaches
  - Inference of sensitive information
  - Training of intrusive models

- Holders of sensitive data cannot benefit from large-scale deep learning because they may not share their datasets for training
  - Biomedical researchers?
  - Social scientists?

Our Objective



Privacy-Preserving Training

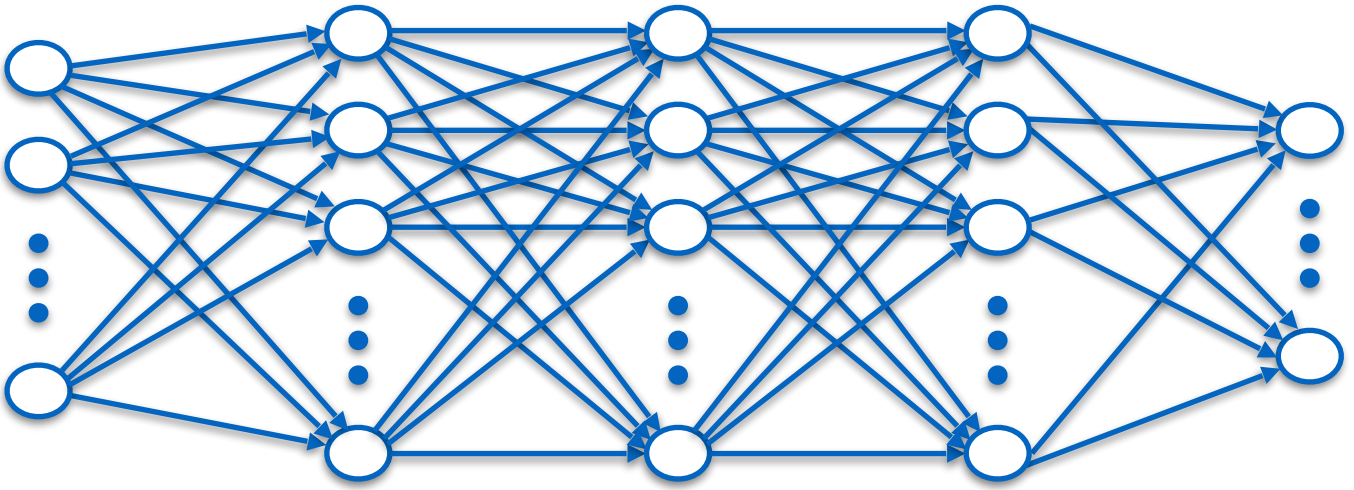


Private data remain private

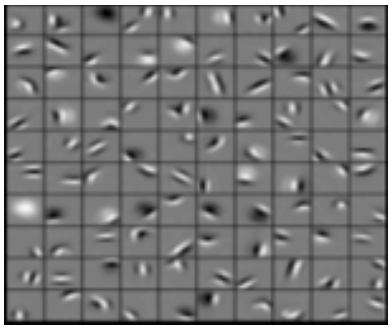
Background

Deep Neural Networks

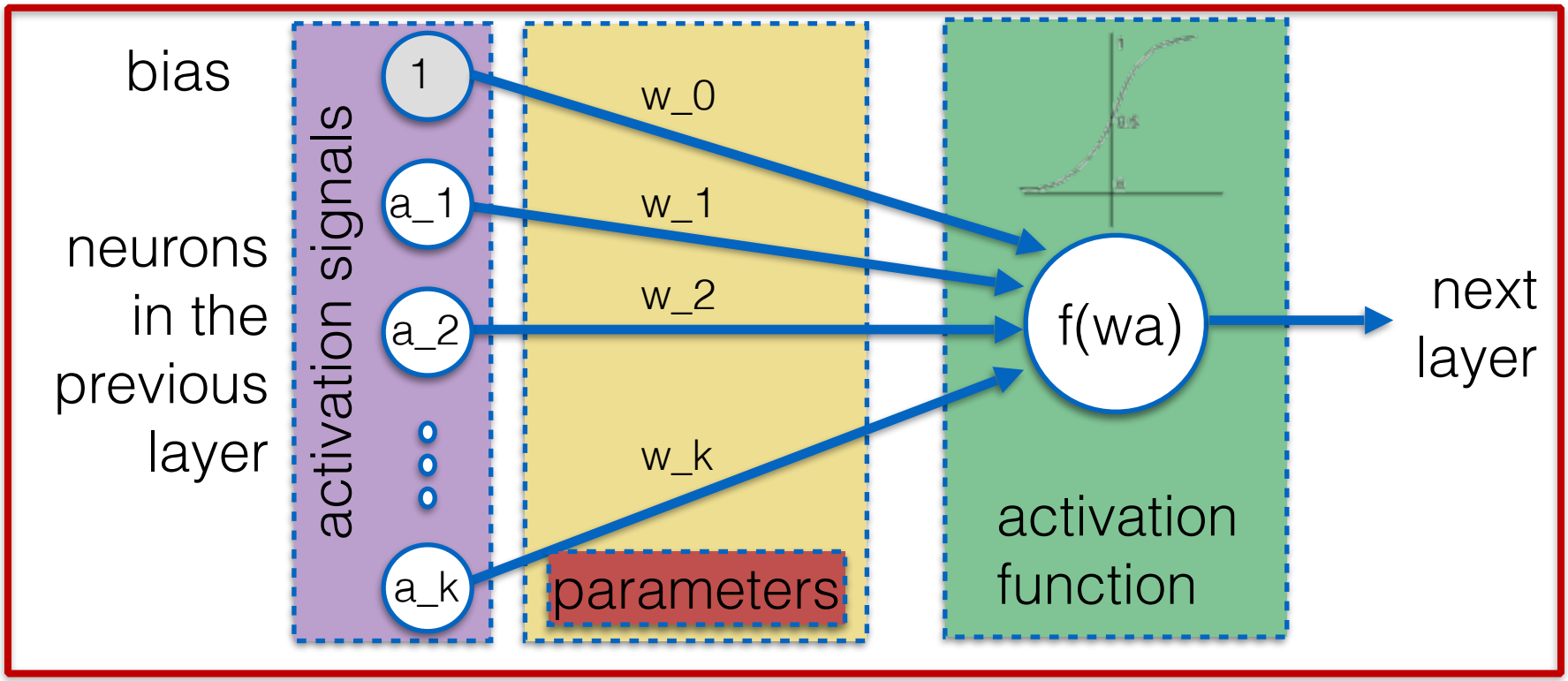
input



output

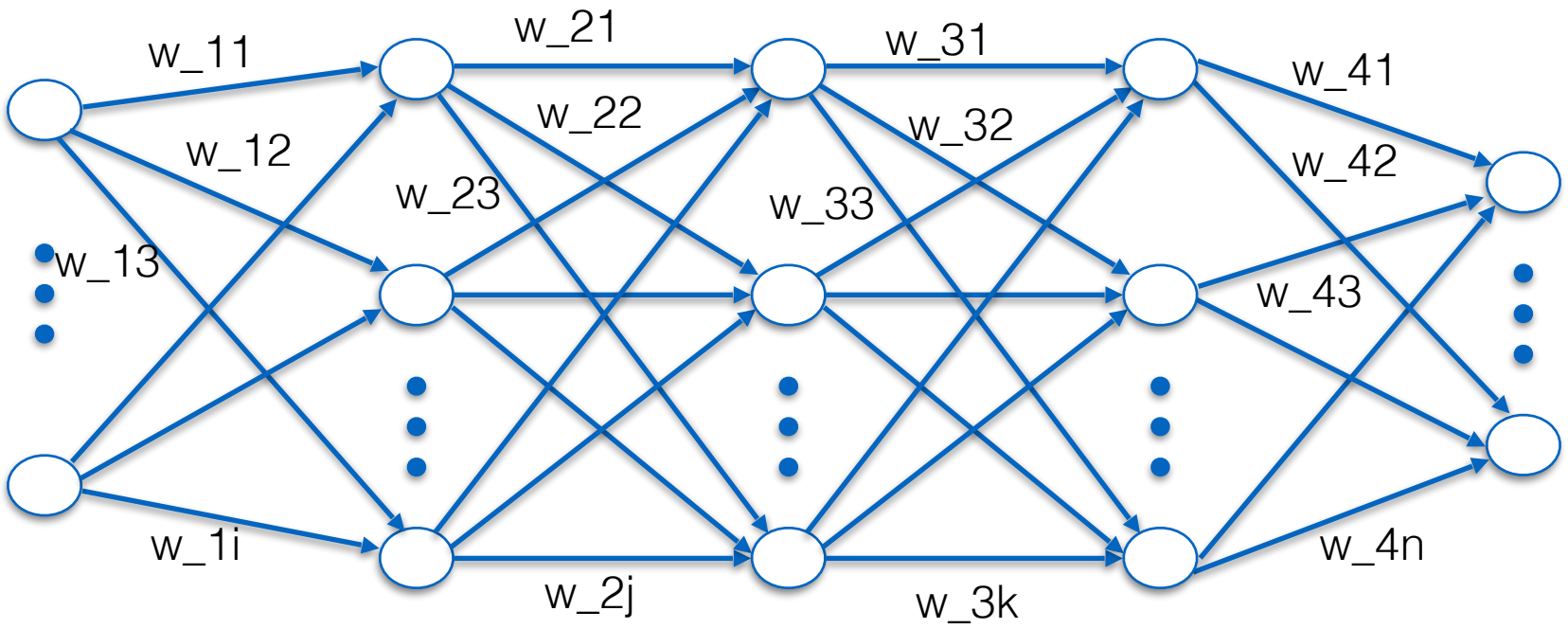


Deep Neural Networks



Learn parameters using
Stochastic Gradient Descent (SGD) algorithm

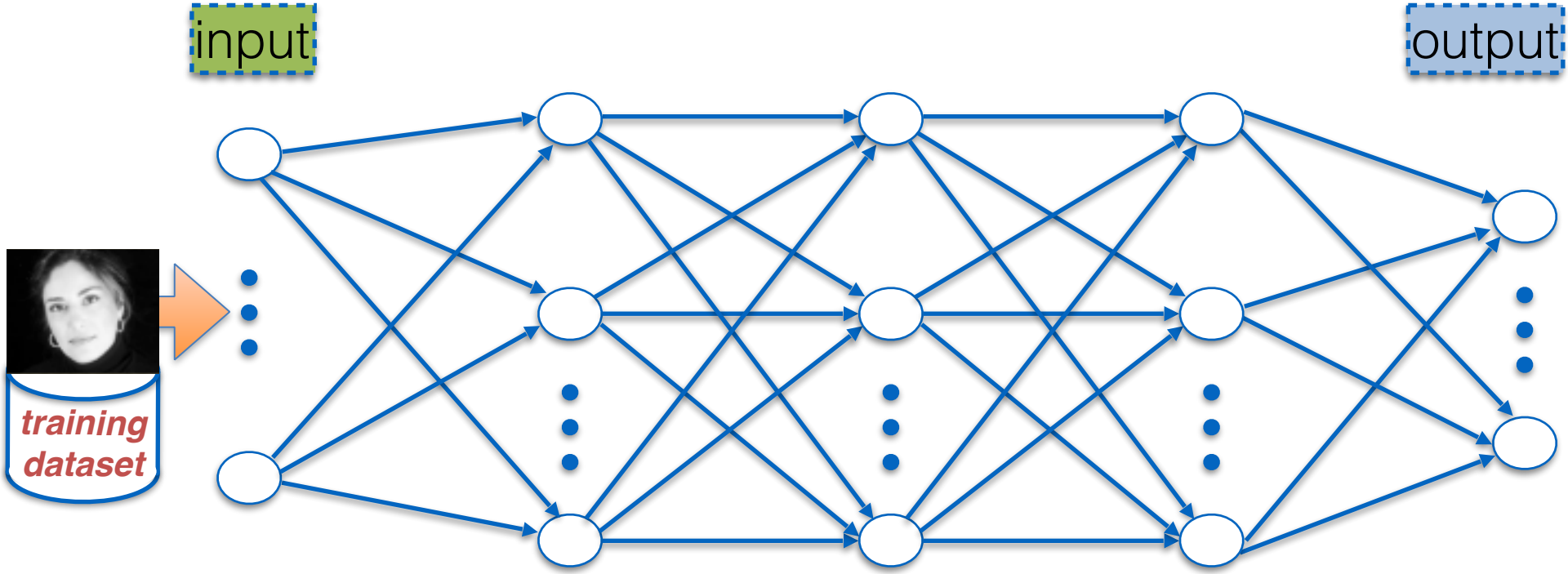
Parameter Training using SGD



- Find parameters that minimize the classification error

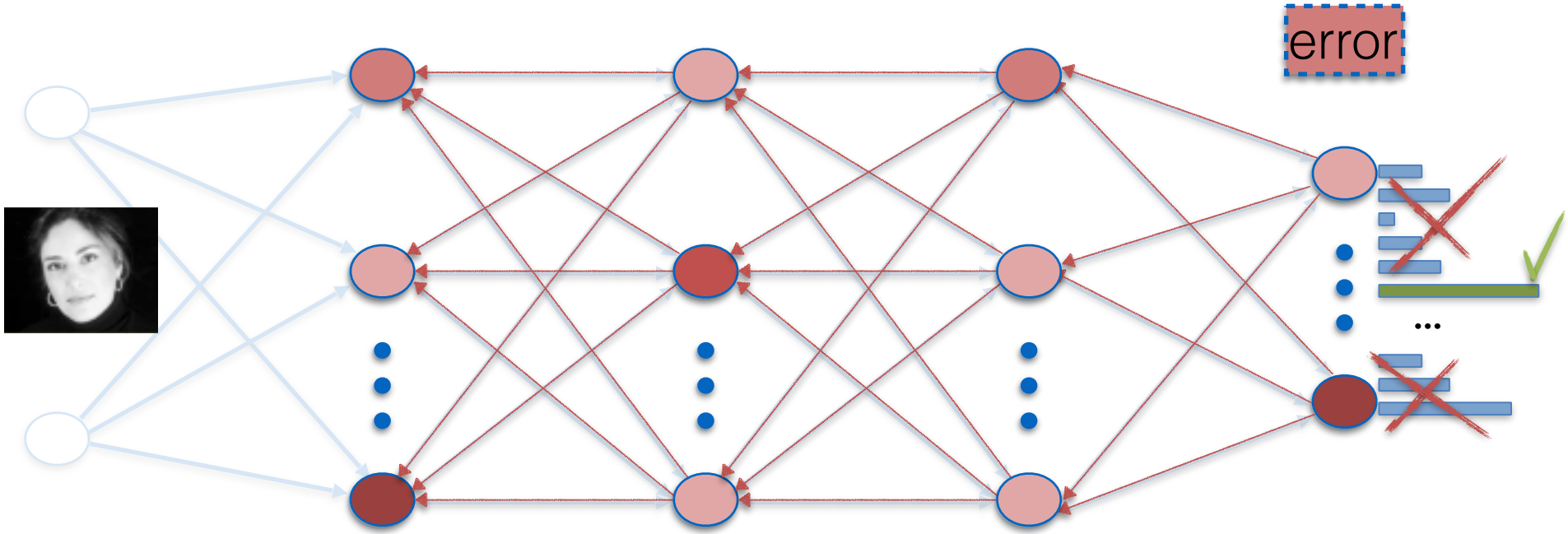
Parameter Training using SGD

1) Feed-Forward

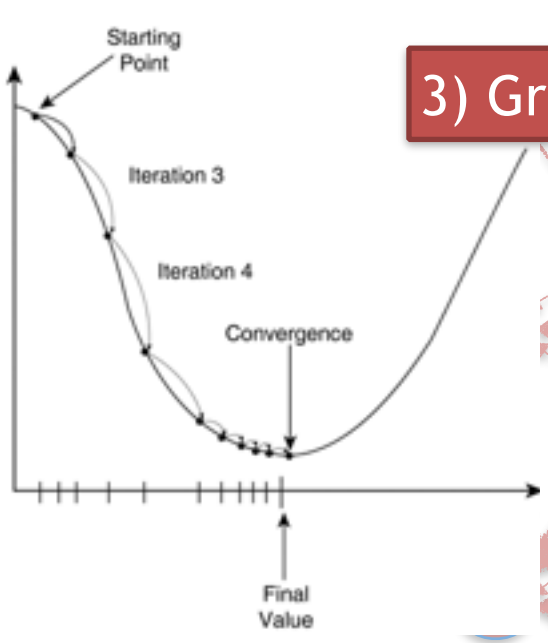


Parameter Training using SGD

2) Back-propagation



Parameter Training using SGD



3) Gradient Descent

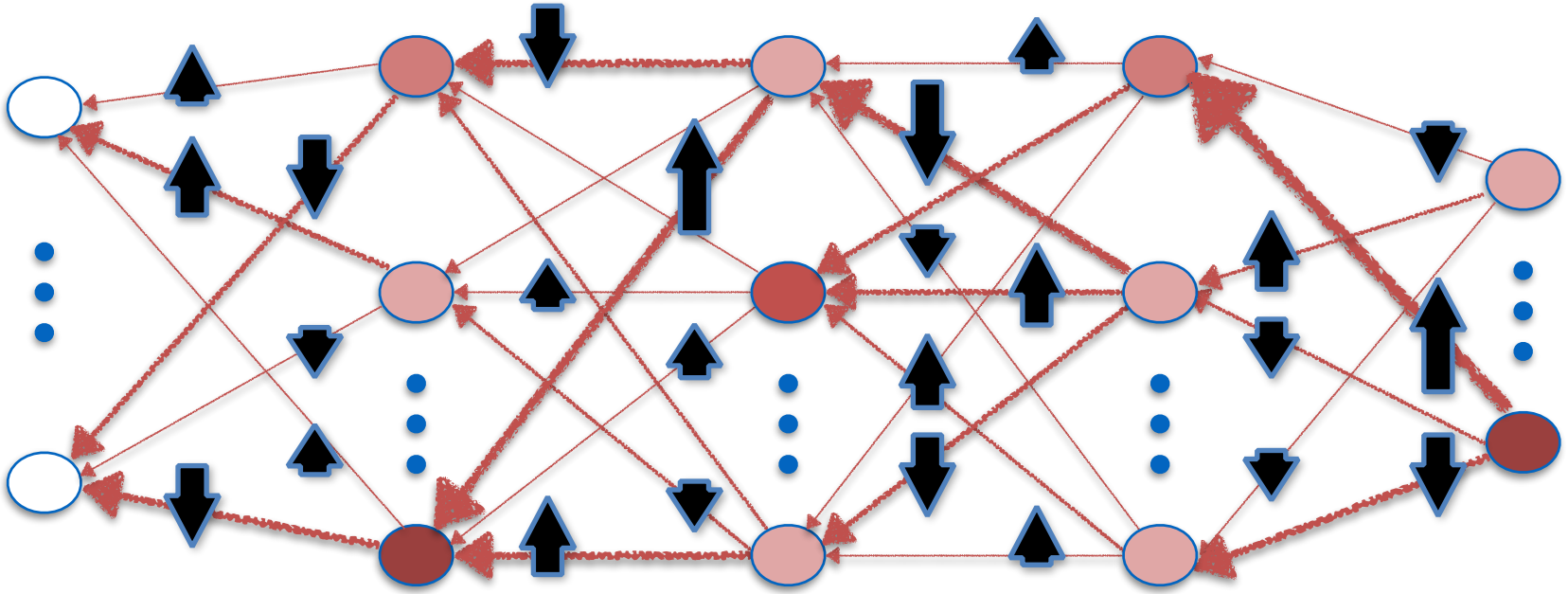
$$w_j := w_j - \alpha \frac{\partial E}{\partial w_j}$$

error E

$$\frac{\partial E}{\partial w_j}$$

Parameter Training using SGD

Parameter Update

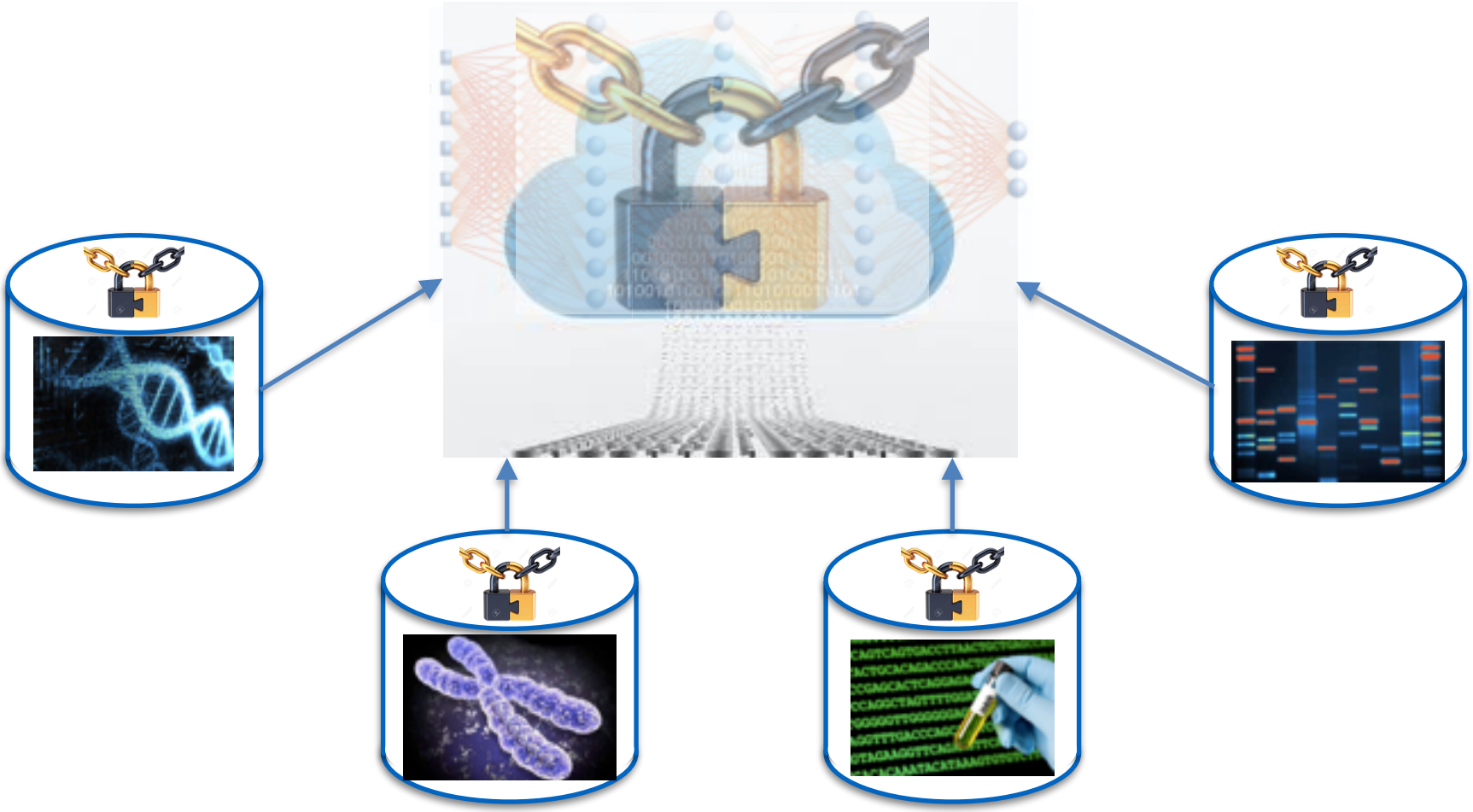


Repeat for new batches of training data

Privacy-Preserving Deep Learning

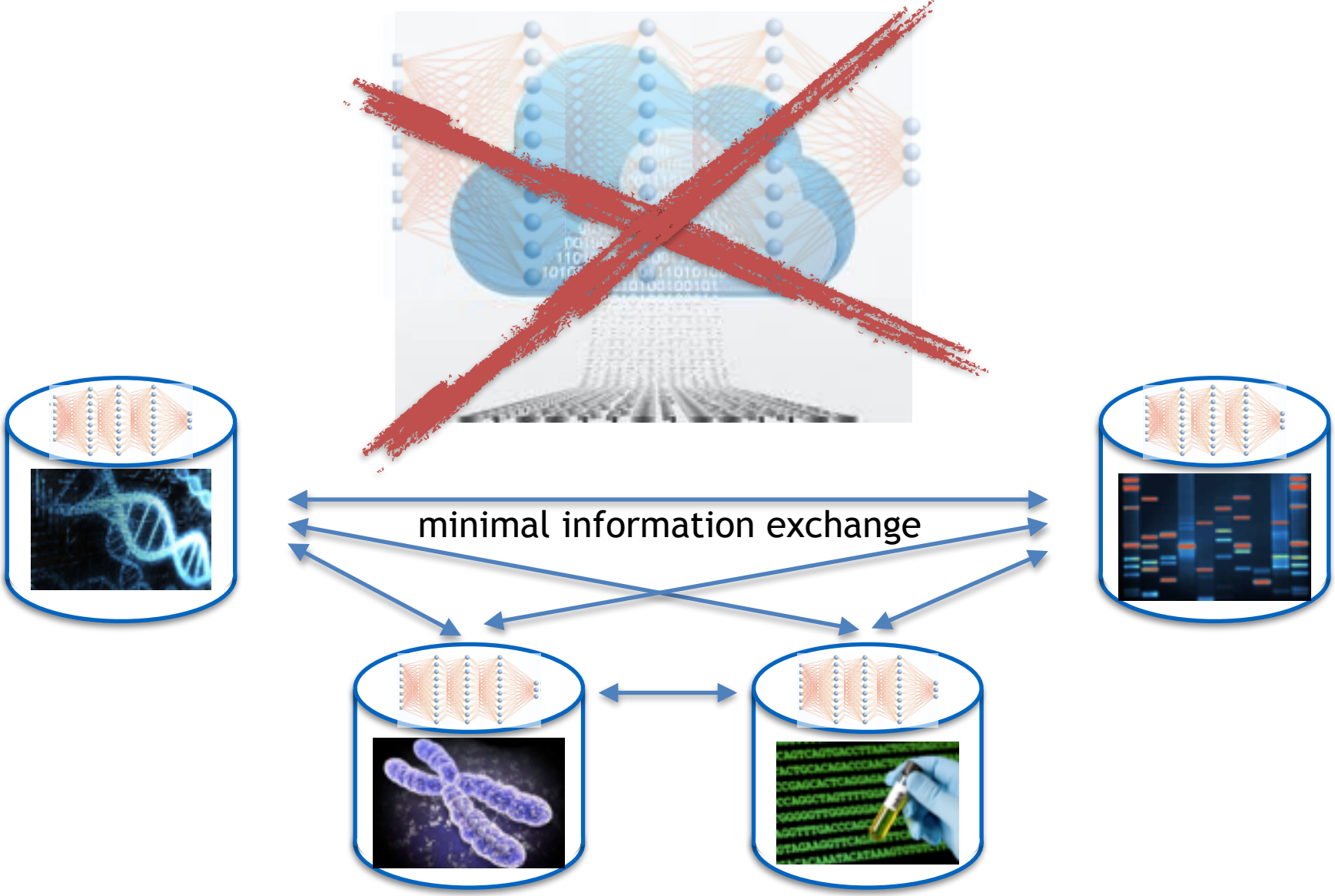
Prior Work

Secure multi-party computation



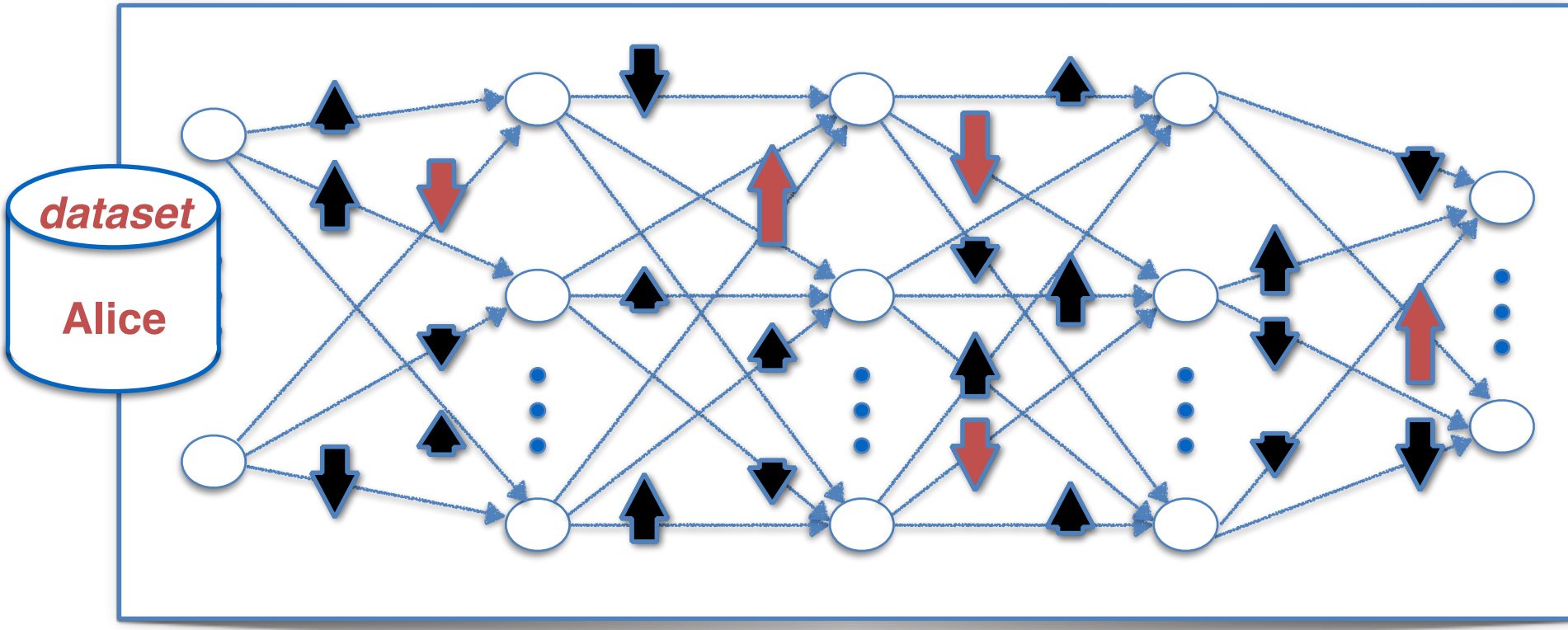
J. Yuan, S. Yu, "Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing", IEEE TPDS, 2014.

A. Bansal, T. Chen, S. Zhong, "Privacy preserving Back-propagation neural network learning over arbitrarily partitioned data", Neural Computing and Applications, 2011



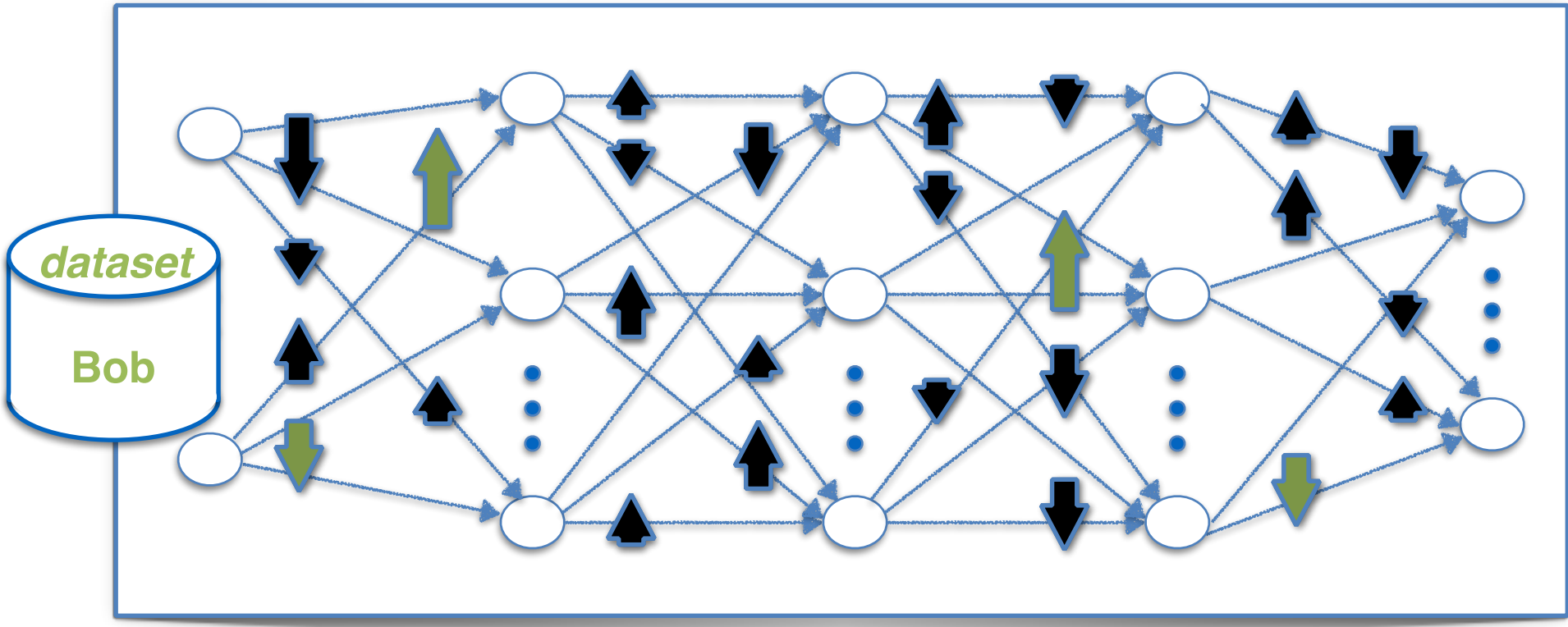
Distributed Selective SGD (DSSGD)

Selective SGD

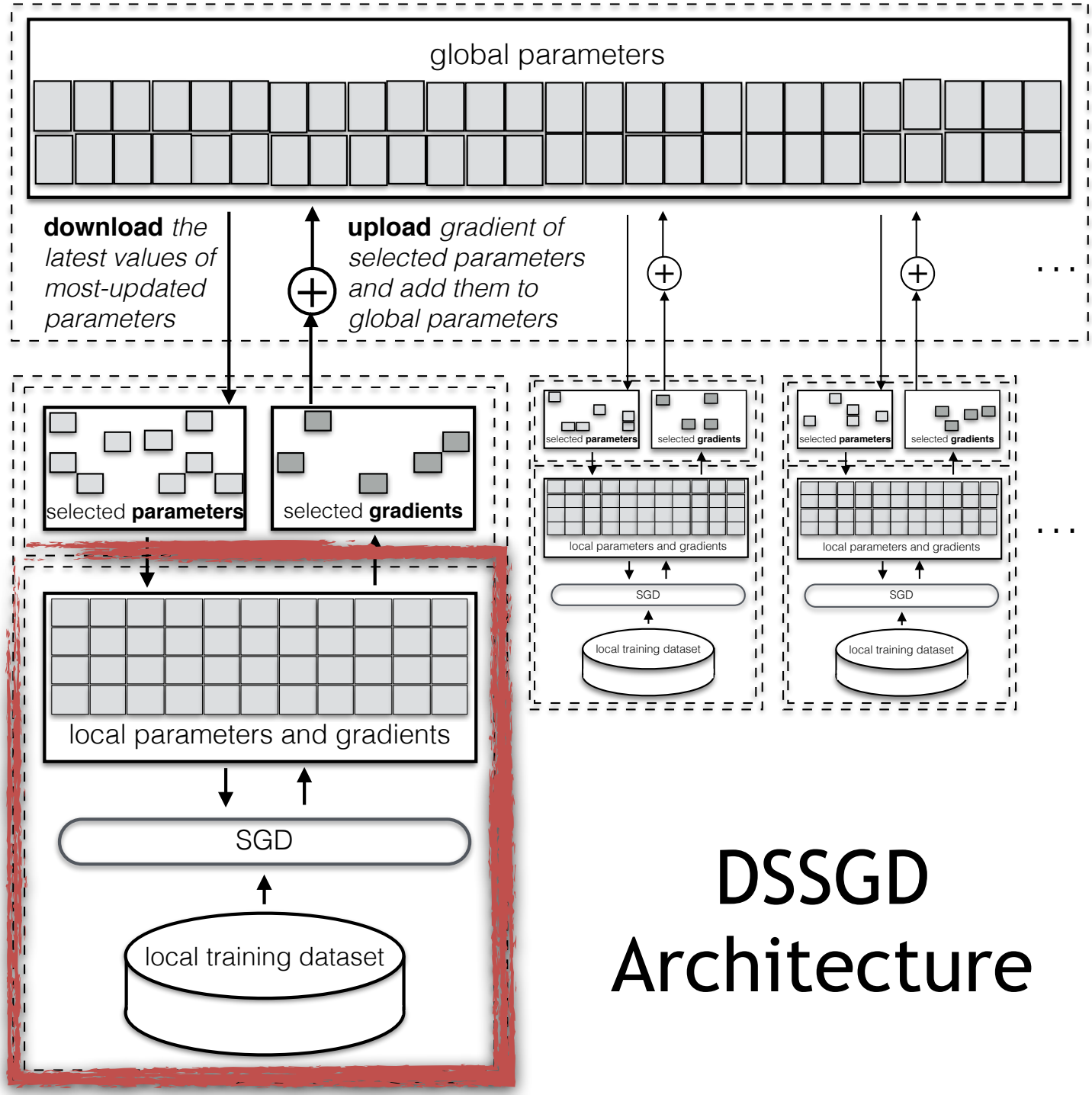


Share with others

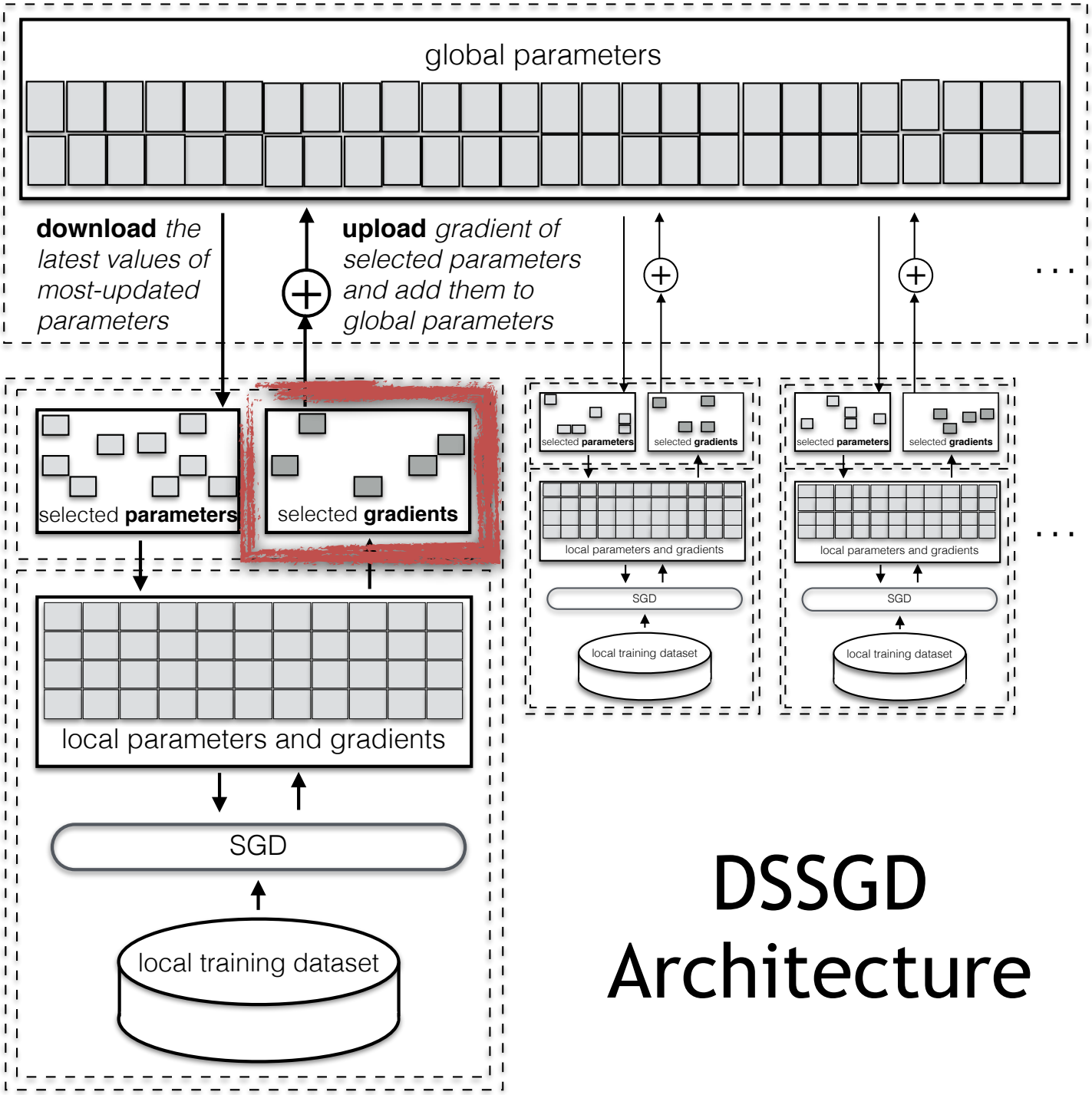
Selective SGD



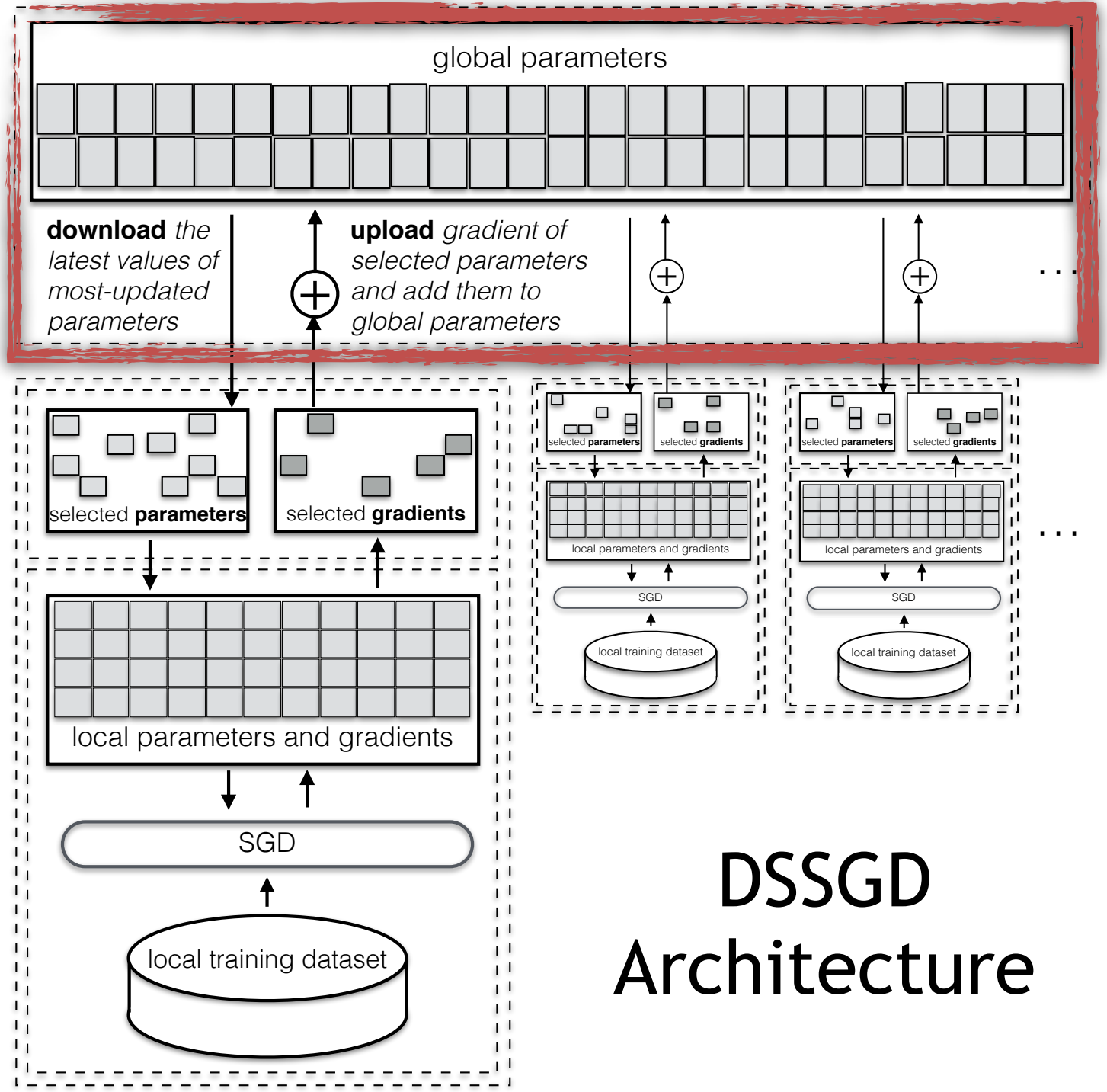
Share with others



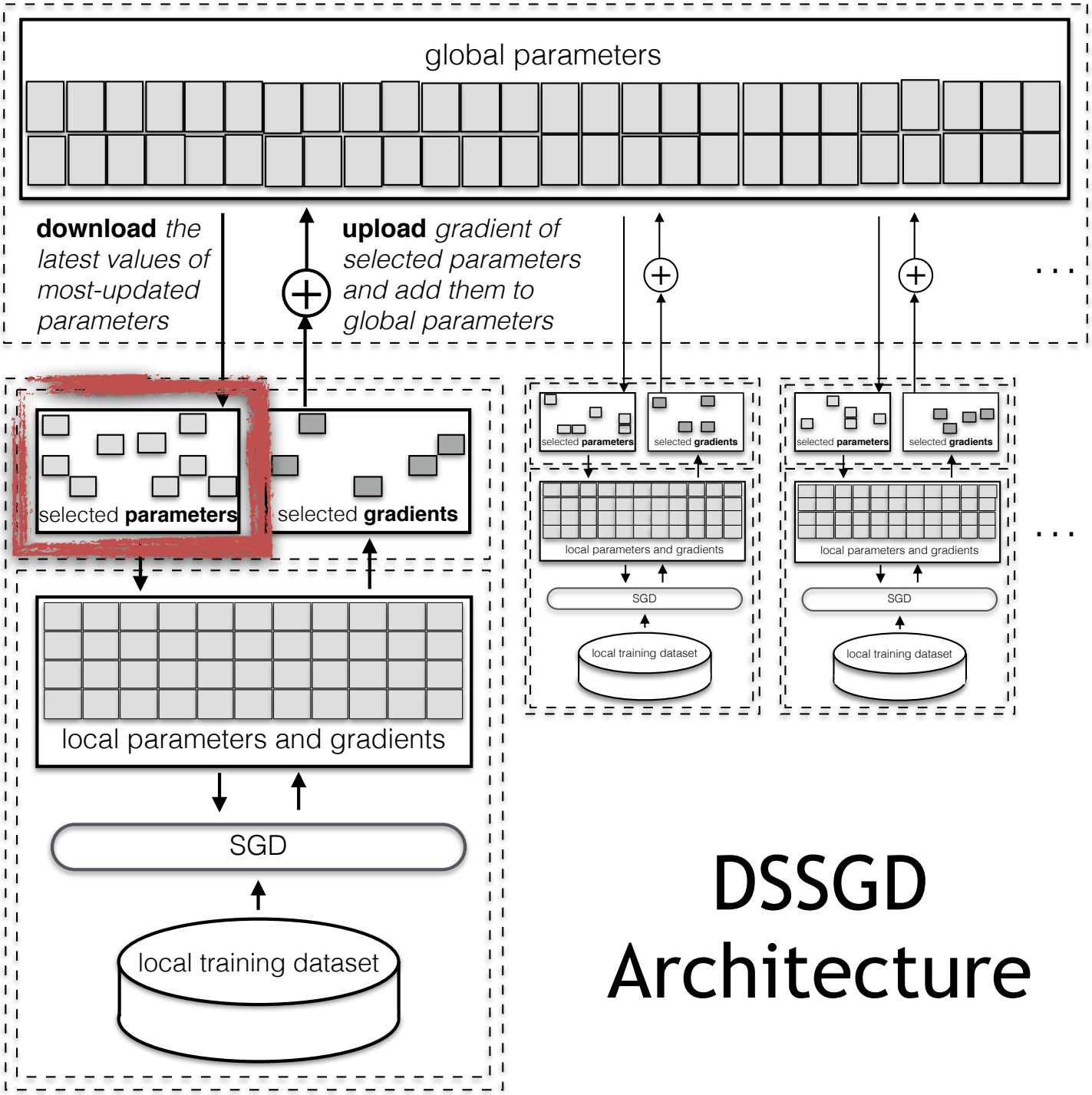
DSSGD Architecture



DSSGD Architecture



DSSGD Architecture



DSSGD Architecture

Distributed Selective SGD

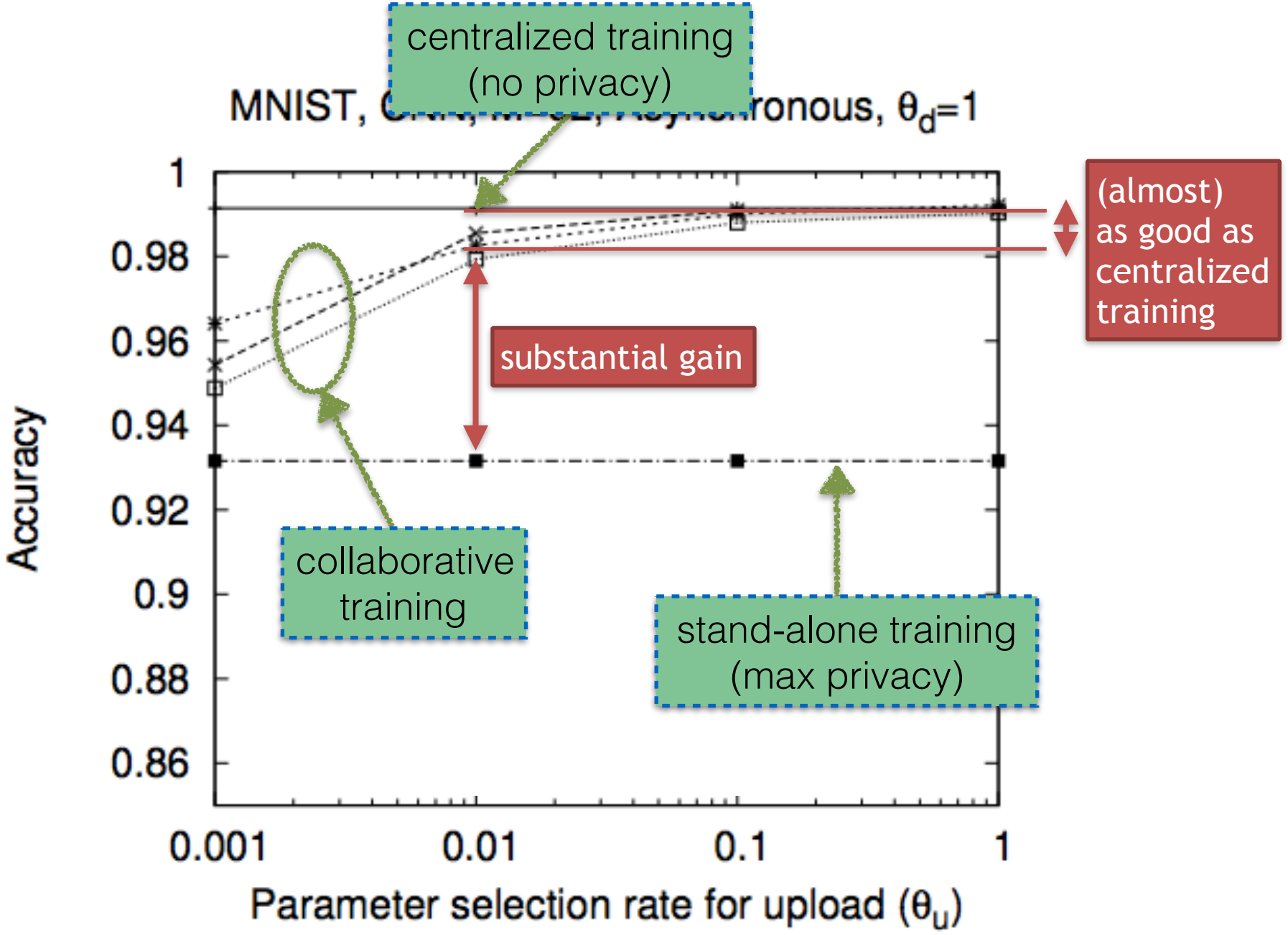
- Local training, Global convergence
- High training stochasticity
- Less overfitting

Evaluation Datasets

MNIST

SVHN

Task: Find the digit in the image (classify into one of 10 classes)

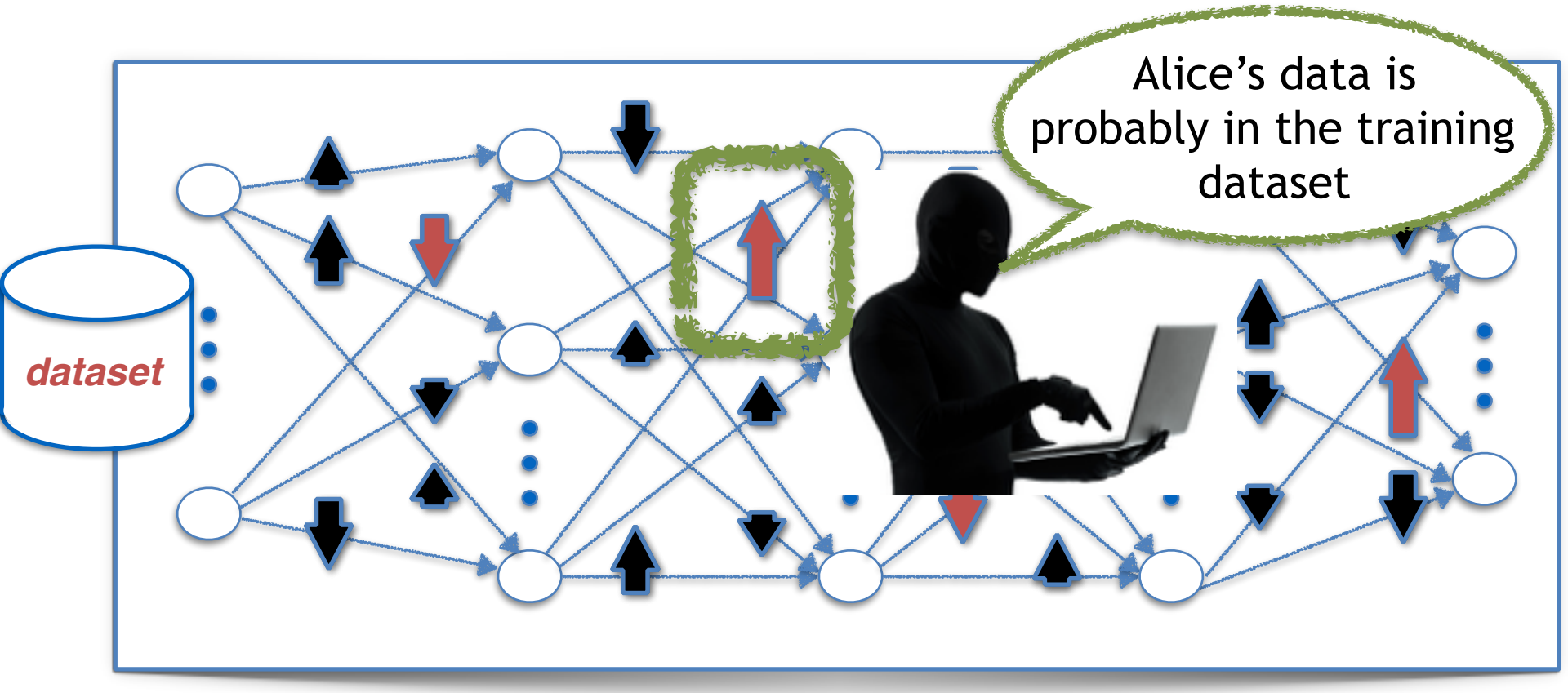

centralized training
(no privacy)
MNIST, ronous, θ_d=1
(almost)
as good as
centralized
training
substantial gain
collaborative
training
stand-alone training
(max privacy)
Accuracy
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.001
0.01
0.1
1
Parameter selection rate for upload (θ_u)

Privacy Properties

- Participants' datasets remain private
- Full control over parameter selection
- Known learning objective
- Resulting model available to all parties

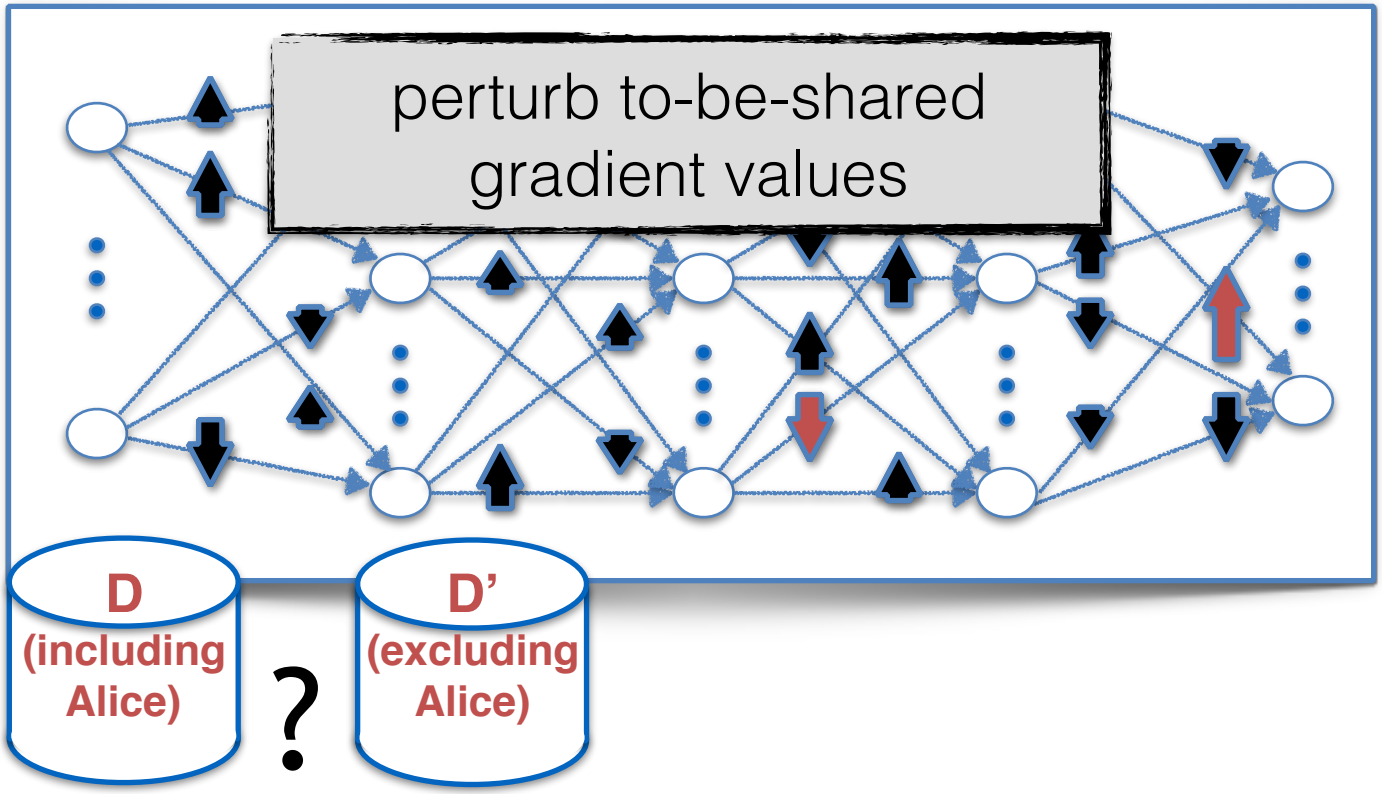
Indirect Information Leakage

through gradient sharing



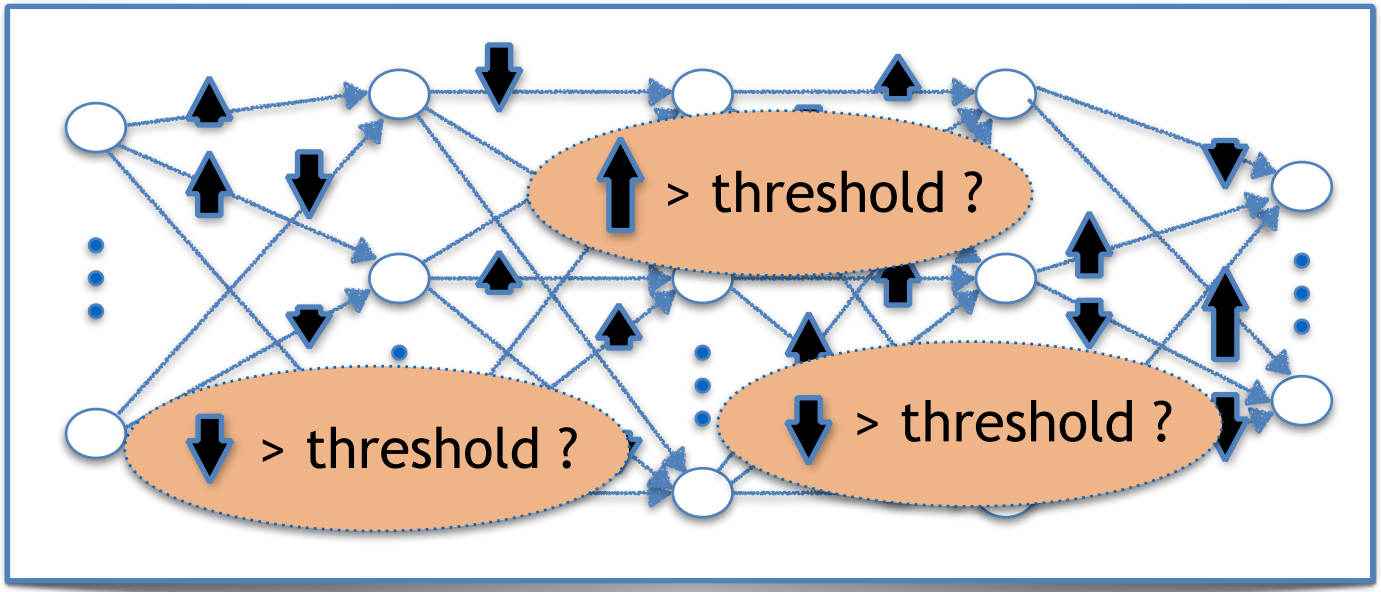
Limit Indirect Info. Leakage

- *Differentially Private* parameter selection and gradient sharing



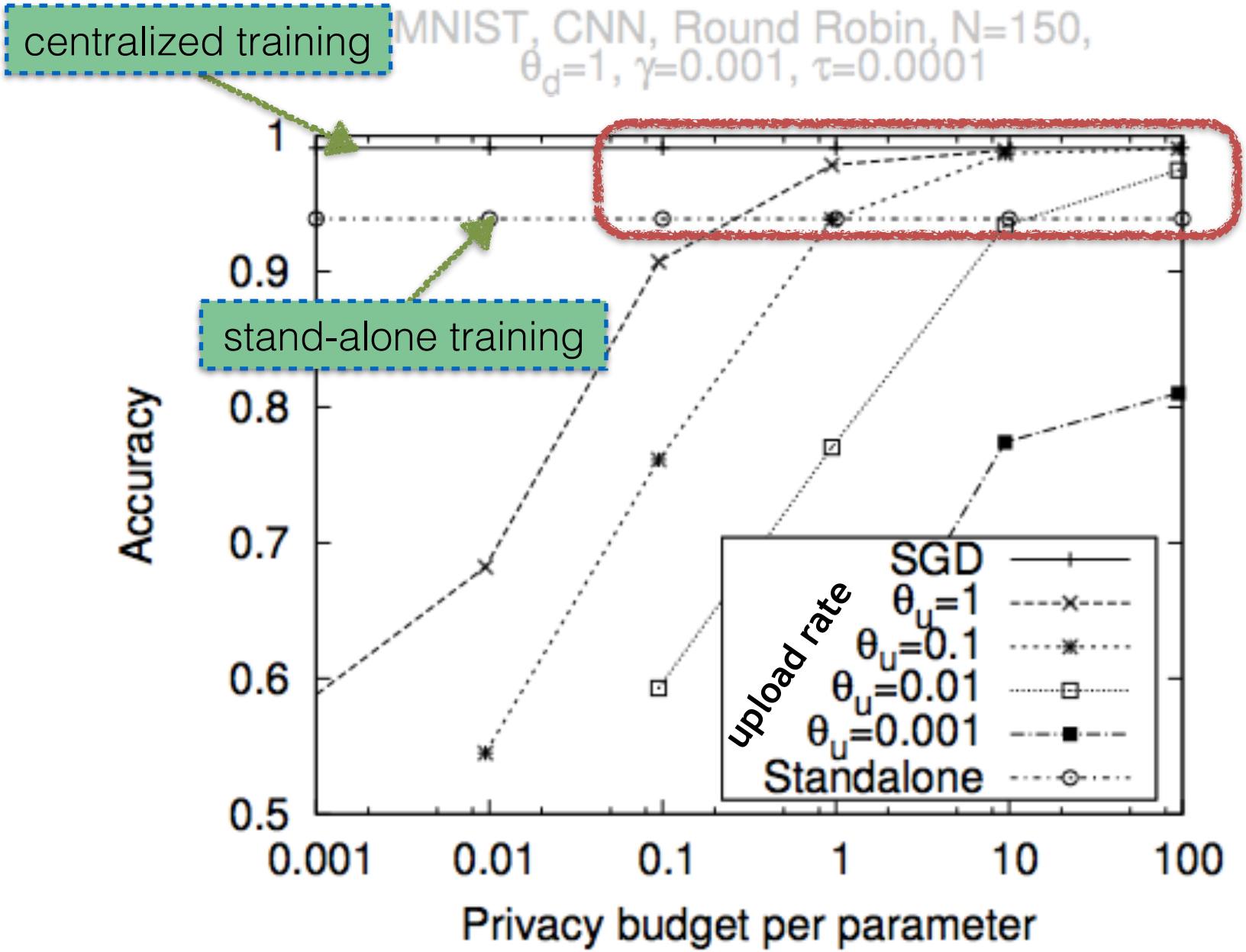
Sparse Vector Technique

- Select a small fraction of (perturbed) gradients that are above a given (perturbed) threshold



C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2013.

differentially-private comparison


centralized training
stand-alone training
MNIST, CNN, Round Robin, N=150,
θd=1, γ=0.001, τ=0.0001
Accuracy
Privacy budget per parameter
1
0.9
0.8
0.7
0.6
0.5
0.001
0.01
0.1
1
10
100
upload rate
SGD
θu=1
θu=0.1
θu=0.01
θu=0.001
Standalone

Conclusions

- Massive data collection required for deep learning presents substantial privacy risks
- We design privacy-preserving deep learning system
  - Holders of sensitive data collaborate to build powerful models without sharing their data
- Differentially-private information exchange protocol
  - Prevents indirect leakage about participants' private datasets
- Accuracy of our system
  - is substantially better than that of standalone learning
  - is very close to accuracy of centralized SGD