ECE 590/COMPSI 590
Special Topics: Edge Computing

Edge for Dispersed Training in Machine Learning

Wednesday October 17th, 2018

Duke UNIVERSITY

Lecture Outline

- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - Federated learning
  - Ensemble learning
  - Challenges and open lines of work
- Personalized ML models with edge

2

Duke UNIVERSITY

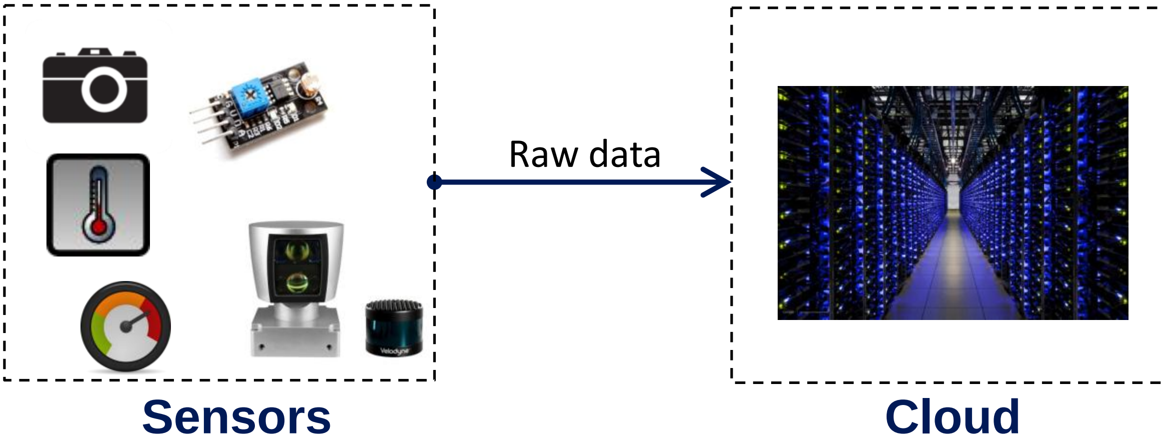
Data Collection Applications

- Slower reaction time
  - ➢Speed of operation less of a concern
  - ➢No in-the-loop control
- Often look across multiple devices

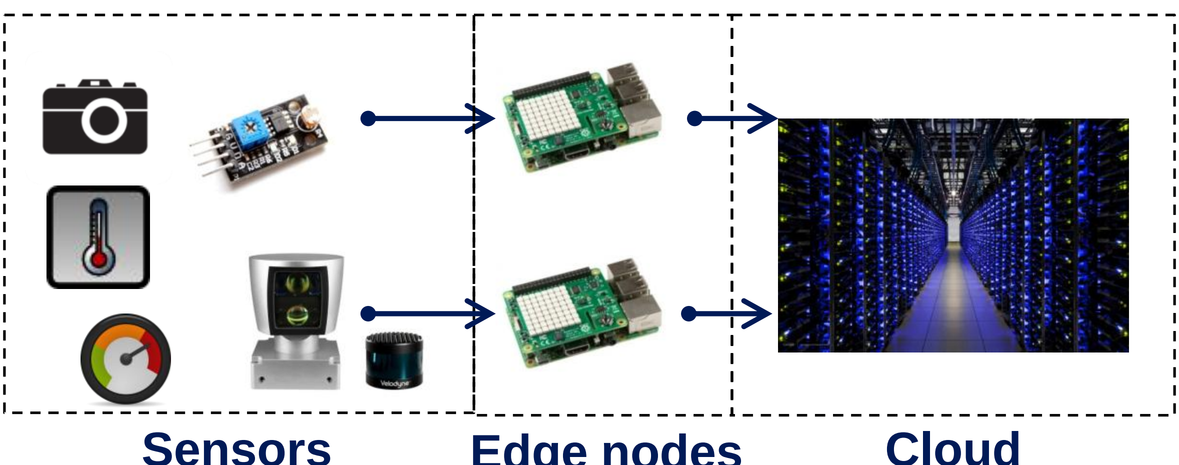
Example Applications: Data Collection Across Multiple Mobile Users

- Images
- Clicks
- System performance
- Lecture reading materials: about these settings

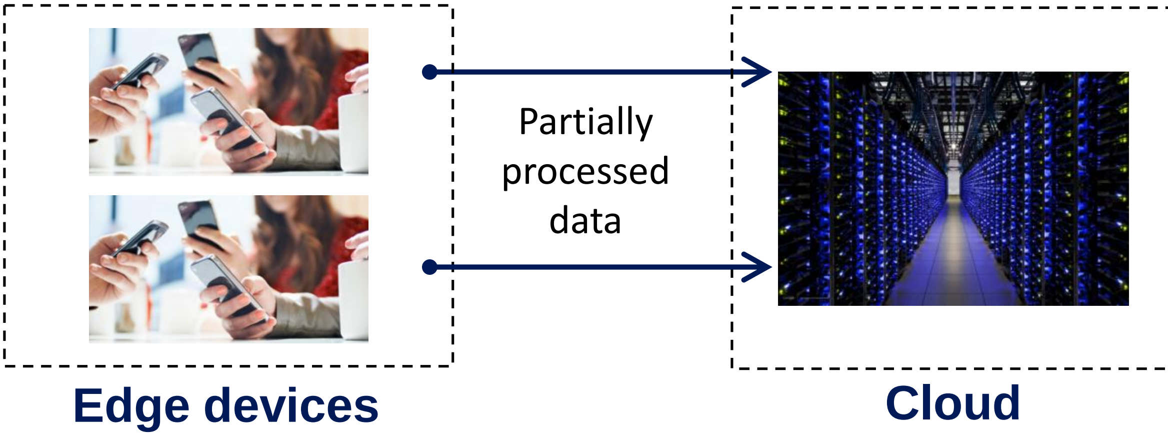
Data Collection Applications: Current State



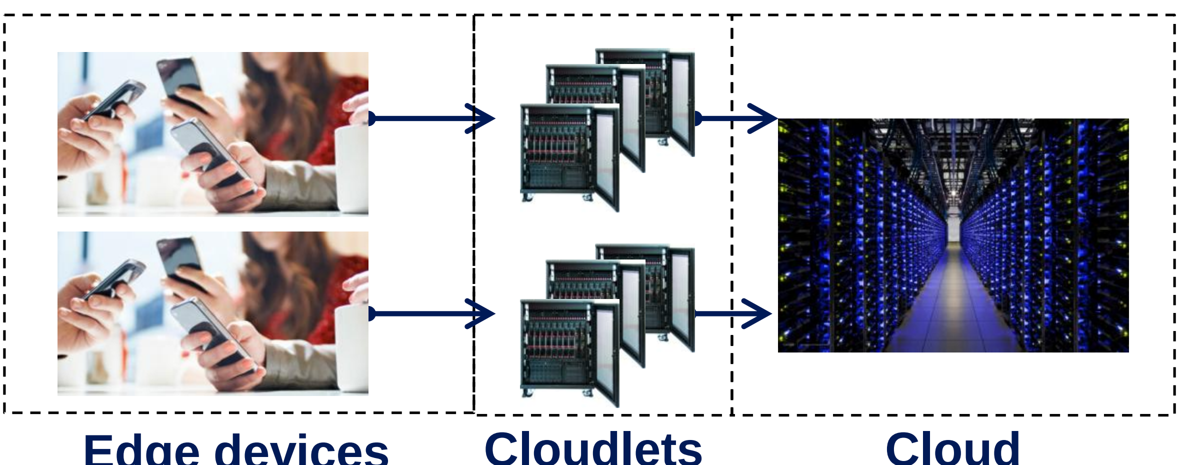
Data Collection Applications: Via Additional Edge Nodes



Data Collection Applications: With Capable Edge Devices



Data Collection Applications: With Edge Devices and Cloudlets



Lecture Outline

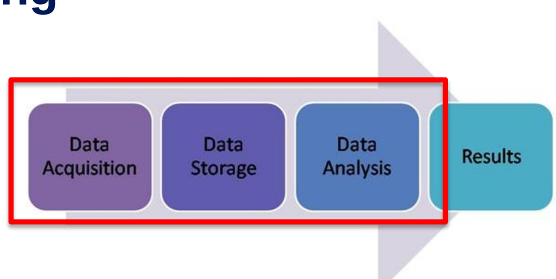
- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - ➢ Federated learning
  - ➢ Ensemble learning
  - ➢ Challenges and open lines of work
- Personalized ML models with edge

9

Duke UNIVERSITY

Data Collection: Comprehensive View

- Change collection approaches and **restructure processing algorithms**
- Requires in-depth understanding of the algorithms
  - ➢ Anomaly detection
  - ➢ **Machine learning algorithms**



10

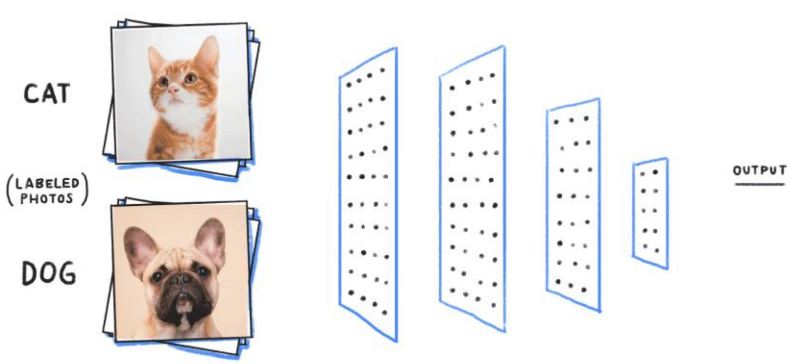
Duke UNIVERSITY

Machine Learning (ML): An Introduction

- Got its name in 1959
- Uses statistical techniques to give systems ability to learn from data

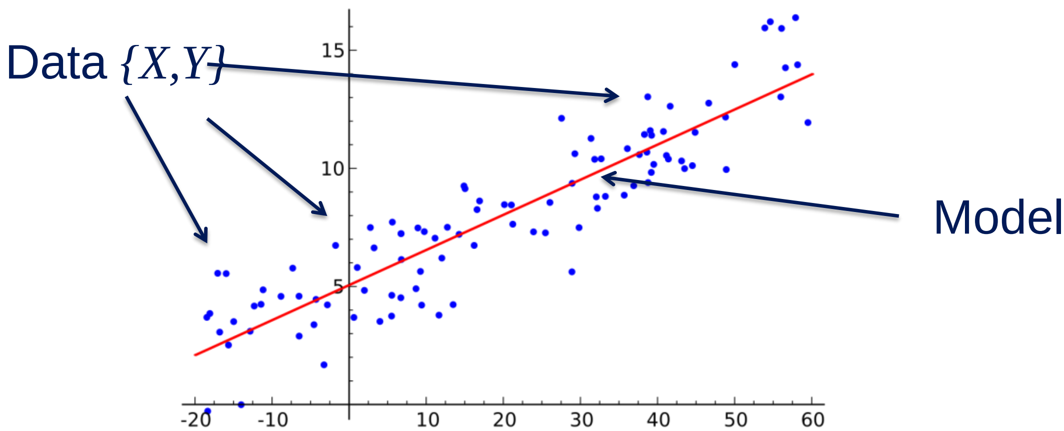
- ML examined in context of this lecture: massive amounts of *collected data* translated into models

ML Application: Classification



Data + labels Model

ML Application: Regression



Traditional Approach to Model Training

- Collect data from multiple devices
- Train ***the*** model
- Send the model to multiple devices

Traditional Distributed Learning

- Collect data from multiple devices
- **Train the model**
  - ➢ *Distributing processing between multiple cores and multiple servers*
- Send the model to multiple devices

Changes Possible With Edge

- Instead of collecting data from multiple devices
  - ➢ Keep some or all data local
- Instead of training ***the*** model
  - ➢ Train different local models: incremental models, private models, …

Emerging lines of work

Lecture Outline

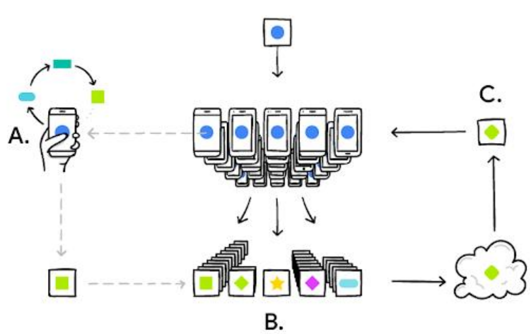
- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - Federated learning
  - Ensemble learning
  - Challenges and open lines of work
- Personalized ML models with edge

17

Duke UNIVERSITY

Keeping (more) Data Local with Edge

- Some of the training happens on the edge device

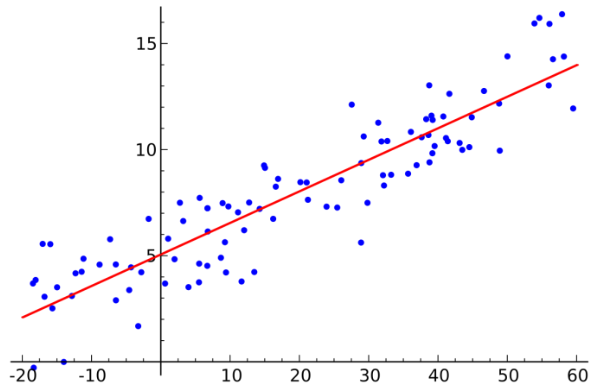


18

Duke UNIVERSITY

Machine Learning as an Optimization Problem

- Fitting a model to data requires minimization of a cost function over all data points
  - ➢ Basis of learning is an **optimization operation**



Stochastic Gradient Descent

- Multiple optimization approaches are possible, but "*recent multitude of deep learning have almost exclusively relied on variants of stochastic gradient descent*"*
- **Stochastic Gradient Descent (SGD):** we take batches of data at a time, uniformly at random, and use the results to move closer to the global minimum

*Communication-Efficient Learning of Deep Networks from Decentralized Data, McMahan et al, arXiv, Feb. 2017.

Federated Learning

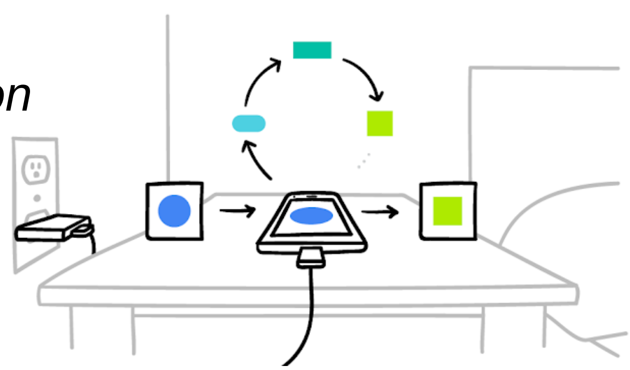
- Take batches of clients at a time
  - Have each client in the batch compute the gradient on its local data, and iterate on it
  - Average the results on the cloud
- Keep going

"Federated learning can be made practical"

Communication-Efficient Learning of Deep Networks from Decentralized Data, McMahan et al, arXiv, Feb. 2017.

Phones Participating Overnight

- *"When the device is idle, plugged in, and on a free wireless connection"*

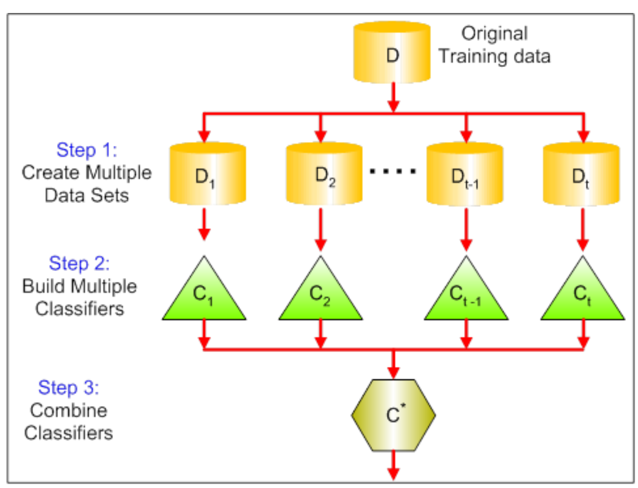


Lecture Outline

- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - ➢ Federated learning
  - ➢ Ensemble learning
  - ➢ Challenges and open lines of work
- Personalized ML models with edge

Ensemble Learning

- Motivation: ML models are often made up of *ensembles* of weaker "*base*" models
  - ➢ KDD-Cup, Netflix Prize are usually won by ensemble-based models

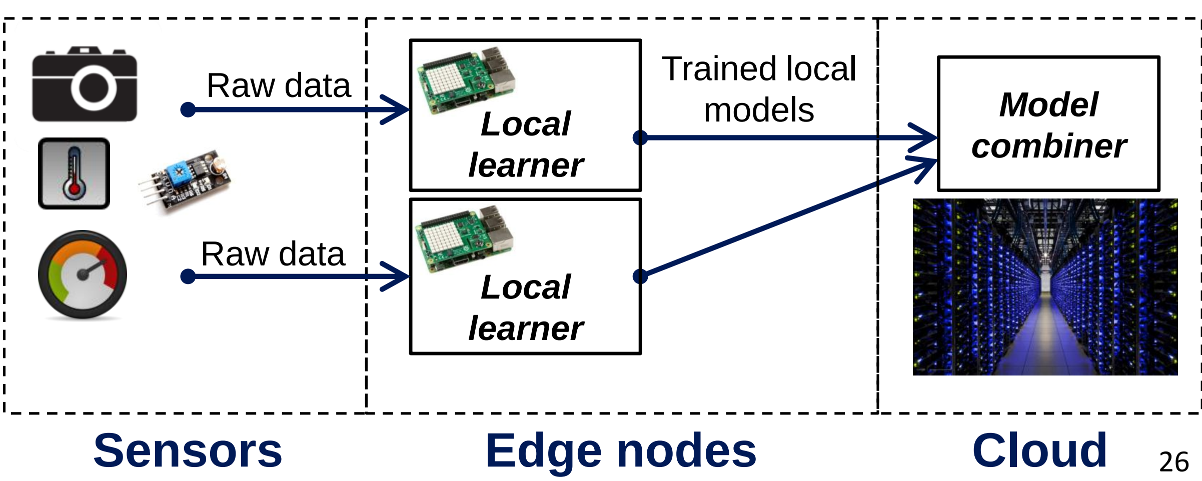


Many Ensemble Learning Techniques Exist

- E.g., bagging: weaker models *voting* in classification tasks, *averaging* in regression tasks
- Heterogeneous and homogeneous base learners
- Base learners generated sequentially or in parallel

In edge systems, can we use locally trained models to form a cloud-based ensemble model?

Ensemble Learning at Edge



Departure from the State of the Art in ML, but Starting Points Exist

- Communication advantage of the approach: *model parameters << raw data*

- Internet-scale ensembles identified as a blue-sky vision before in ensemble learning

Ensemble Methods: Foundations and Algorithms, Zhi-Hua Zhou, Chapman&Hall, 2012

27

Duke UNIVERSITY

Our Early-Stage Work: Ensemble Learning on the Edge

- *Adaptive Regression by Mixing* (ARM) to combine locally computed models
  - ➢ Train local models
  - ➢ Send some raw data to the cloud as well
  - ➢ Combine models with the help of raw data
- Achieves substantial bandwidth reduction

- ❑ Y. Ruan, L. Zheng, M. Gorlatova, M. Chiang, C. Joe-Wong, The Economics of Fog Computing: Pricing Tradeoffs for Distributed Data Analytics, *Fognet and Fogonomics, Wiley*, in print, 2019.
- ❑ T. Chang, L. Zheng, M. Gorlatova, C. Gitau, C. Huang, M. Chiang, Demo: Decomposing Data Analytics in Fog Networks, *ACM SenSys'17*, Delft, Netherlands, Nov. 2017.

Lecture Outline

- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - ➢ Federated learning
  - ➢ Ensemble learning
  - ➢ Challenges and open lines of work
- Personalized ML models with edge

29

Duke UNIVERSITY

Statistical Challenges (1/2)

- Traditional learning algorithms assume that data is I.I.D.
  - ➢ Reasonable assumption for large datasets
  - ➢ SGD, some ensemble learning methods pick data points at random to ensure I.I.D. properties
- Assumption **fundamental** to convergence analysis and other "nice" properties

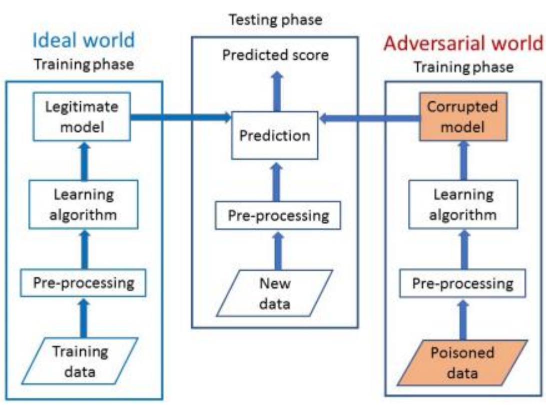
30

Duke UNIVERSITY

Statistical Challenges (2/2)

- Local data: non-I.I.D.
- And also:
  - Of uneven size
  - With unbalanced class representations

Security Challenges



- Local models could be easier to poison

Image from: Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning, arXiv, Apr. 2018

System Challenges

- Storage, computation, communication, energy restrictions
  - ➢ Not considered by designers of machine learning algorithms

Many opportunities for research

Not as Clear of a Use Case as Inference on the Edge (1/3)

- Inference on the edge delivers important performance improvements
- Performance-wise, ML *training* on the edge is no better than on the cloud
  - ➢ Local devices are not as capable as cloud-scale systems
  - ➢ Latency is not a factor
  - ➢ Quality can be slightly worse
  - ➢ No new applications are enabled by it

Not as Clear of a Use Case as Inference on the Edge (2/3)

- From the development perspective, having data in one place is useful
- Can process it as you'd like, whenever you'd like
  - Parameter tuning
  - Retraining
  - Understanding when more data could be needed
- Use case assumptions seem to clash with down-to-earth software/ML development practices

Not as Clear of a Use Case as Inference on the Edge (3/3)

- Bandwidth is saved
  - Philosophical question: are we inventing a very complex *lossy compression* mechanism?
- Privacy-preservation is an advantage
  - But is this true privacy?

Several Lines of Current Work

- Federated Multi-Task Learning – NIPS'17
  - ➢ Fitting separate but related models simultaneously, in federated settings
- Adaptive Federated Learning – IEEE INFOCOM'18
  - ➢ Deciding on an appropriate degree of federation

Lecture Outline

- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - ➢ Federated learning
  - ➢ Ensemble learning
  - ➢ Challenges and open lines of work
- Personalized ML models with edge

Personalized Training

- Why have **the** one model for each and every set of conditions?
  - Opportunities for: person-specific and location-specific performance

Why look for a cat in a video feed from a location without a cat?

Why look for a generic dog, rather than this specific one?

Model Specialization in MCDNN

- Look for inputs that are *(n,p)-dominated*
  - *n* most frequent classes account for at least *p* fractions of the weights
- Add specialized models to model catalog
- A somewhat brute-forced online approach
  - "*Typically many of the specialized variants are strictly worse than others*"

From: MCDNN: An Approximation-Based Execution Framework for Deep Stream Processing Under Resource Constraints, Han et al, ACM MobiSys'16.

Personalized Training: New Paradigm for Machine Learning

- Supervised learning long focused on creating the one best model from data

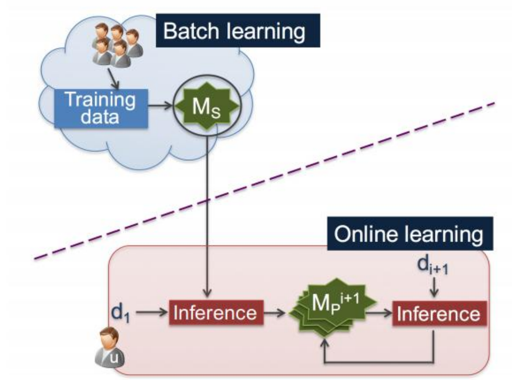
- In edge computing: privacy as an additional motivating factor

Fully Private Training: On Local Data Alone

- Not particularly powerful
  - ➢Not enough pictures of cats in restricted local conditions
- A step backwards from the current state of ML

Cloud + Edge Training (1/4)

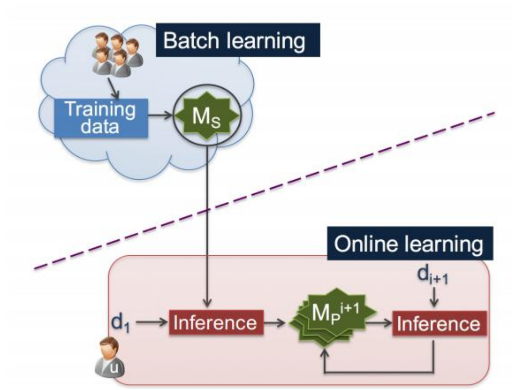
- Local incremental training of a globally generated shared model



From: Privacy-preserving Personal Model Training, Servia-Rodriguez et al, IEEE/ACM IoTDI'18.

Cloud + Edge Training (2/4)

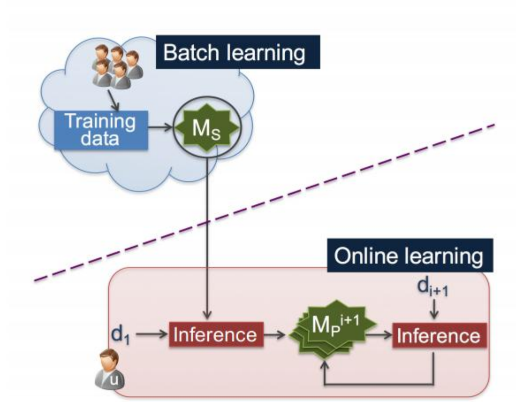
- Device obtains a global model and starts making inferences
- Local data is collected in the meantime



From: Privacy-preserving Personal Model Training, Servia-Rodriguez et al, IEEE/ACM IoTDI'18.

Cloud + Edge Training (3/4)

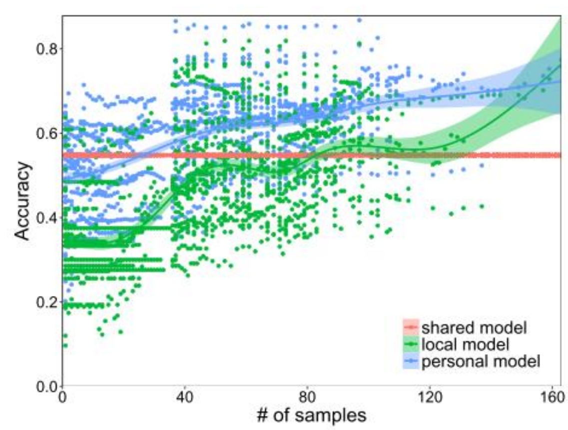
- *Personalized* model is generated once there is enough data
  - ➢ Multi-layer Perceptron (MLP) weights, biases updated based on local data



From: Privacy-preserving Personal Model Training, Servia-Rodriguez et al, IEEE/ACM IoTDI'18.

Cloud + Edge Training (4/4)

- High-quality *personalized model* can be generated more readily than a high-quality *local model*



From: Privacy-preserving Personal Model Training, Servia-Rodriguez et al, IEEE/ACM IoTDI'18.

Edge + Edge Training

- Settings: massive amounts of data distributed across multiple locations
  - ➢ Video data
  - ➢ Medical data
- Allowing local classifiers to be different from each other reduces communication overhead

From: Collaborative Probably Approximately Correct (PAC) Learning, Blum et al, NIPS 2017.

Lecture Recap

- Data processing applications in edge computing
- Traditional ML training and ML training involving edge
- Keeping (more) ML data local with edge
  - ➢ Federated learning
  - ➢ Ensemble learning
  - ➢ Challenges and open lines of work
- Personalized ML models with edge