ECE 590/COMPSI 590
Special Topics: Edge Computing

Edge for Dispersed Training in Machine Learning

Wednesday January 29th, 2020

Duke UNIVERSITY

1

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

2

Data Collection Applications

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

3

Duke UNIVERSITY

3

Example Applications: Data Collection Across Multiple Mobile Users

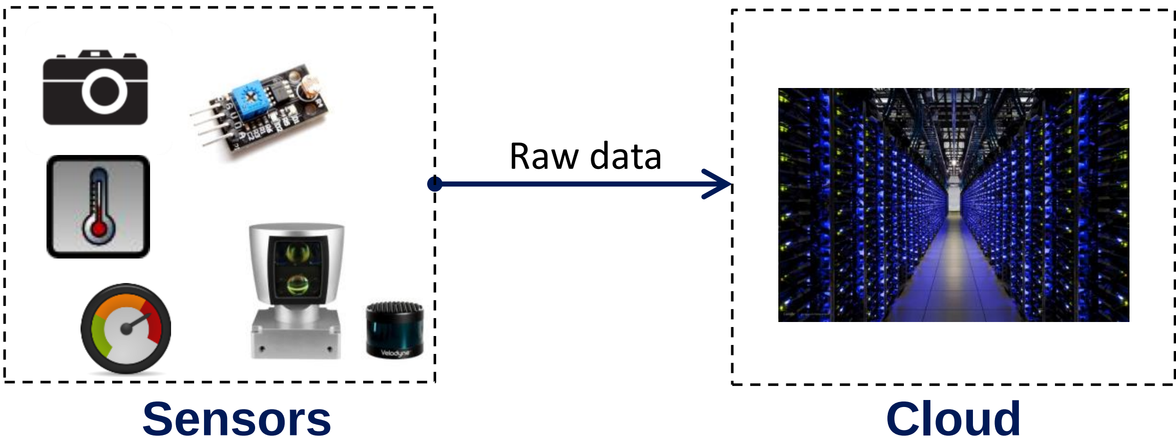
- Images
- Clicks
- System performance

4

Duke UNIVERSITY

4

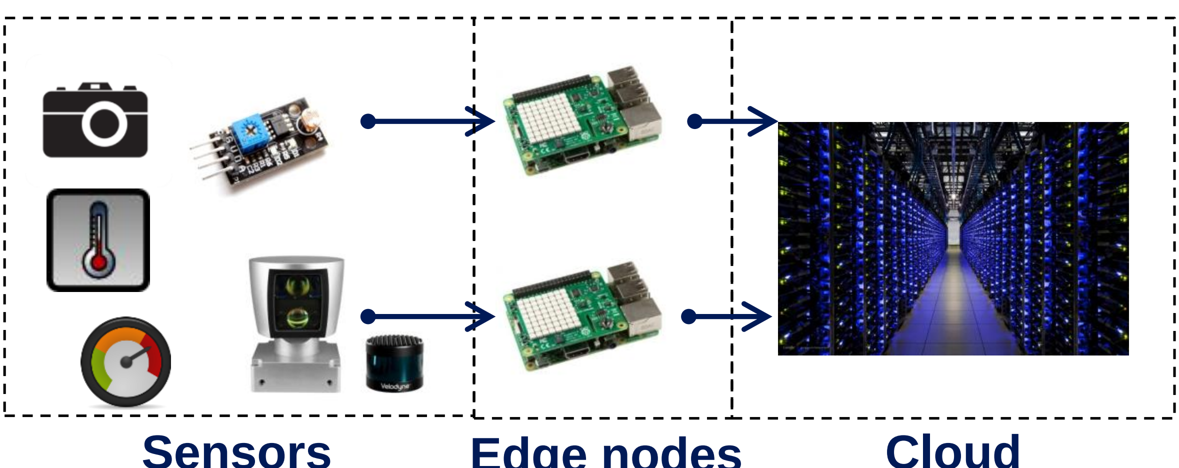
Data Collection Applications: Current State



5

5

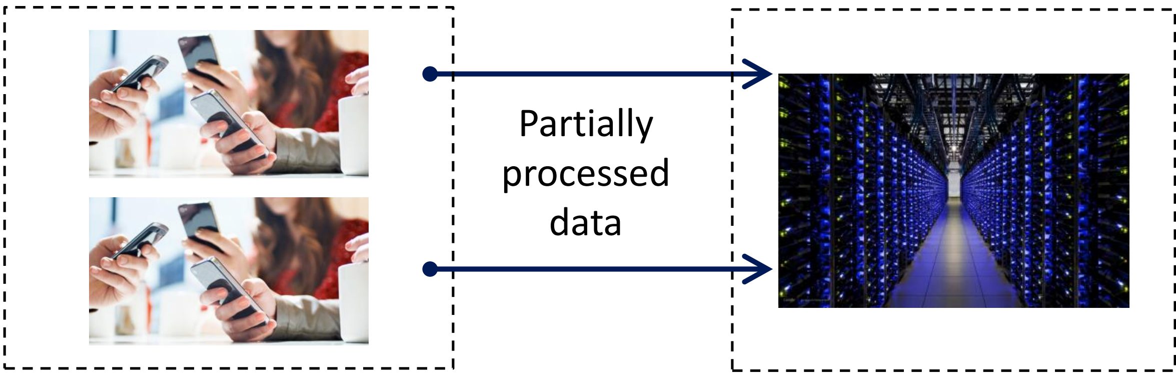
Data Collection Applications: Via Additional Edge Nodes



6

6

Data Collection Applications: With Capable Edge Devices

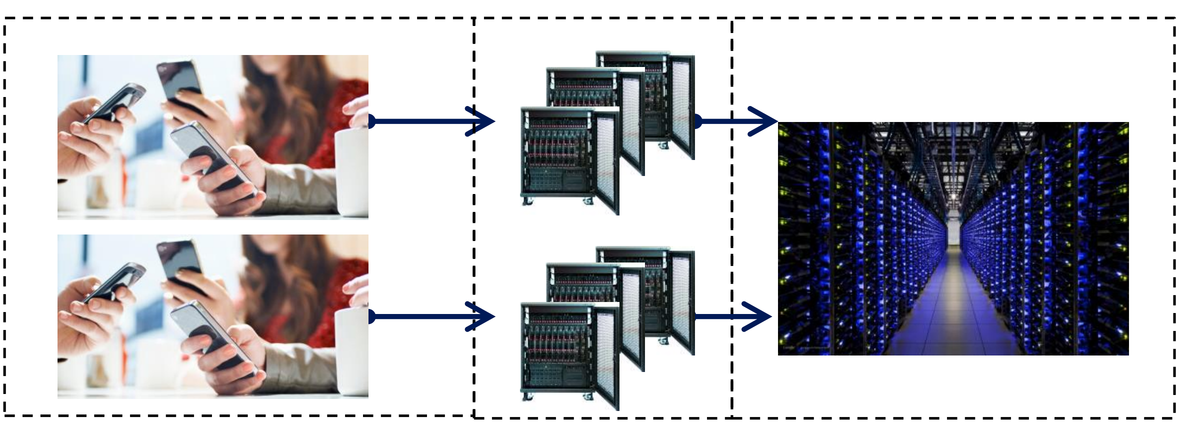


Edge devices **Cloud**

7

7

Data Collection Applications: With Edge Devices and Cloudlets



Edge devices **Cloudlets** **Cloud**

8

8

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

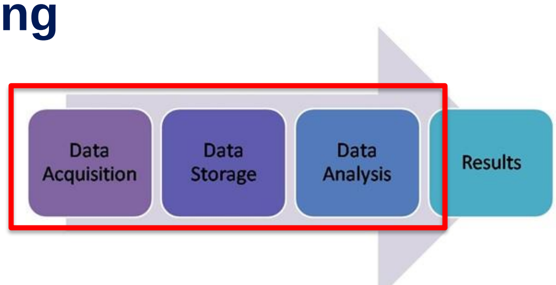
9

Duke UNIVERSITY

9

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

10

Machine Learning (ML): An Introduction

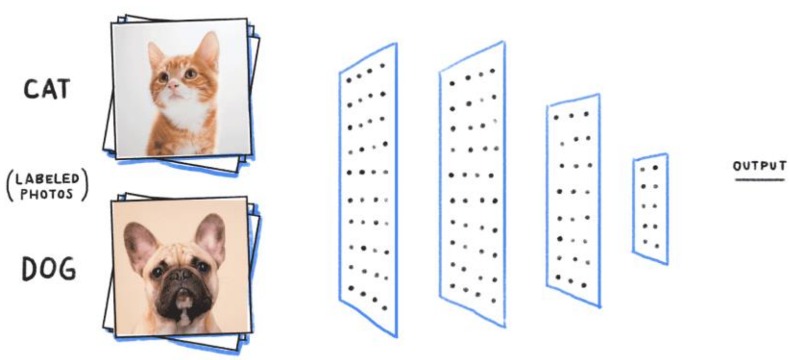
- Uses statistical techniques to give systems ability to learn from data
- ML examined in context of this lecture: massive amounts of *collected data* translated into models

11

Duke UNIVERSITY

11

ML Application: Classification



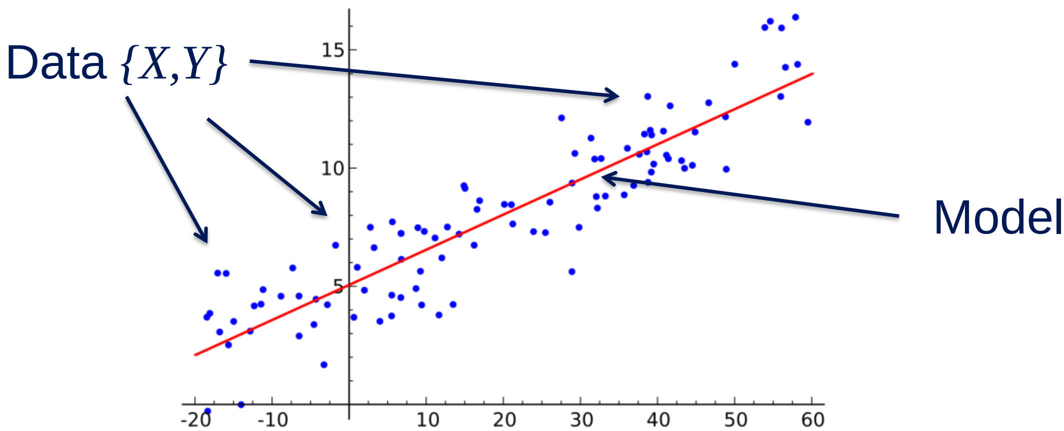
Data + labels Model

12

Duke UNIVERSITY

12

ML Application: Regression



13

Duke UNIVERSITY

13

Traditional Approach to Model Training

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

14

Duke UNIVERSITY

14

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

15

Duke UNIVERSITY

15

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

16

Duke UNIVERSITY

16

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

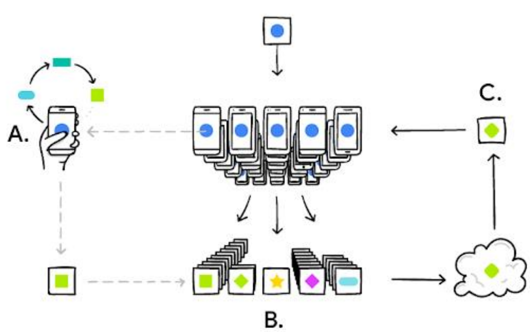
17

Duke UNIVERSITY

17

Keeping (more) Data Local with Edge

- Some of the training happens on the edge device



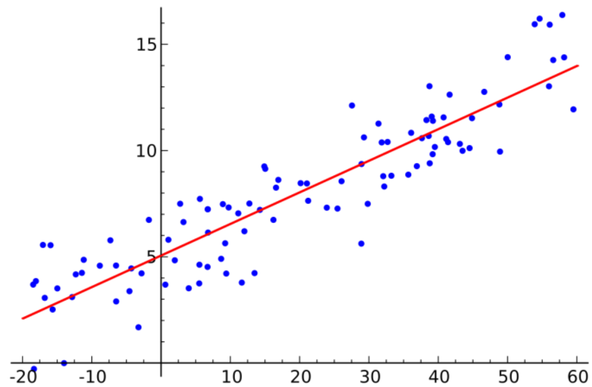
18

Duke UNIVERSITY

18

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**



19

Duke UNIVERSITY

19

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.

20

Duke UNIVERSITY

20

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.

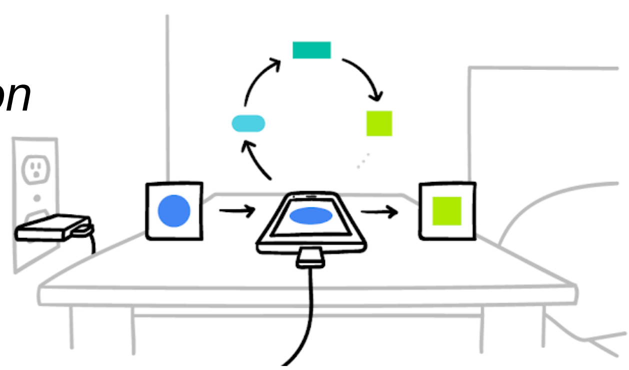
21

Duke UNIVERSITY

21

Phones Participating Overnight

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



22

Duke UNIVERSITY

22

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

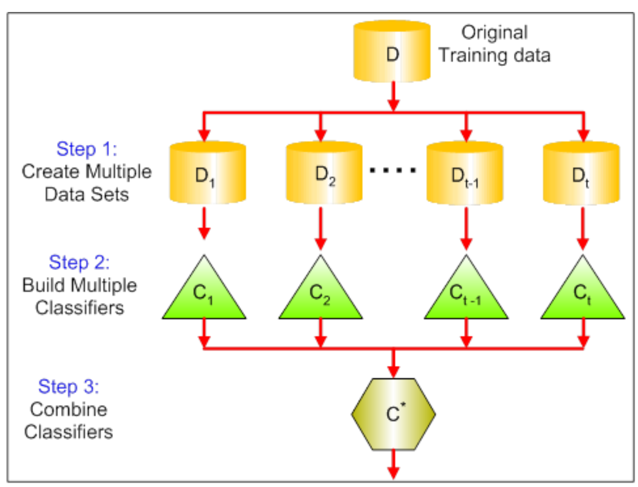
23

Duke UNIVERSITY

23

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



24

Duke UNIVERSITY

24

Many Ensemble Learning Techniques Exist

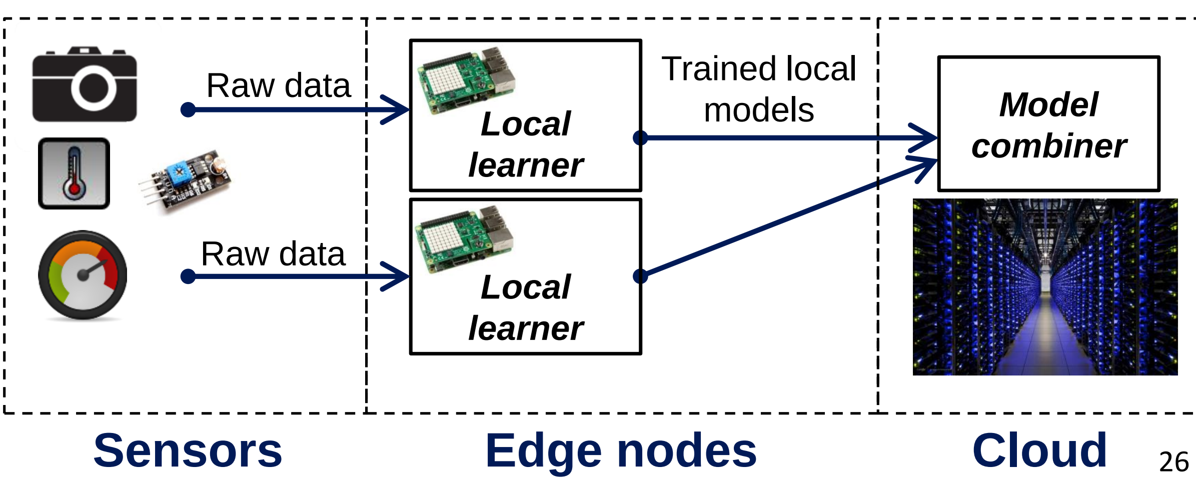
- E.g., bagging: weaker models *voting* in classification tasks, *averaging* in regression tasks
- Heterogeneous and homogeneous base learners
- Base learners generated sequentially or in parallel

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

25

25

Ensemble Learning at Edge



26

26

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

27

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

28

Duke UNIVERSITY

28

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

29

Duke UNIVERSITY

29

Statistical Challenges (2/2)

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

30

Duke UNIVERSITY

30

System Challenges

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

Many opportunities for research

31

Duke UNIVERSITY

31

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

32

Duke UNIVERSITY

32

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

33

Duke UNIVERSITY

33

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?

34

Duke UNIVERSITY

34

Very Popular Research Topic

- Towards federated learning at scale, differentially private federated learning, adaptive federated learning, …
  -

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

Duke UNIVERSITY

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?

37

Duke UNIVERSITY

37

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.

38

Duke UNIVERSITY

38

Personalized Training: New Paradigm for Machine Learning

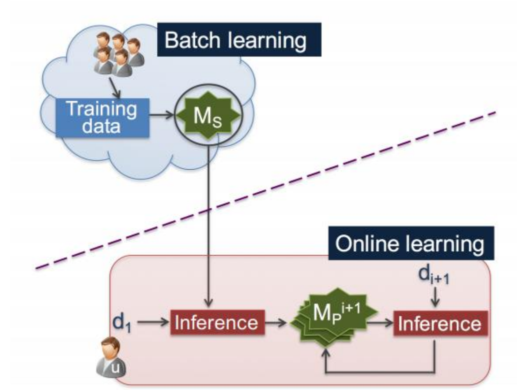
- Supervised learning long focused on creating the one best model from data
- In edge computing: privacy as an additional motivating factor

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

- Local incremental training of a globally generated shared model



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

41

Duke UNIVERSITY

41

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.

42

Duke UNIVERSITY

42

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

43

Duke UNIVERSITY

43

Next Class

- Paper presentations:
  - ➢ Caroline Potts: *Real-time Video Analytics: The Killer App for Edge Computing*
  - ➢ Jiyao Hu: *MUTE: Bringing IoT to Noise Cancellation*

44

Duke UNIVERSITY

44

Homework

- Read the two papers that will be presented
  - Prepare to discuss them; 10 min Q&A for each of the papers
- Work on your research project
  - And specifically your proposal

45

Duke UNIVERSITY

45