ICML 2018 Highlights

Chan Y. Park (ML2 @ KC & Moru Labs)

Experiences

Web & System Programmer (2003 ~ 2005)

B.S. in Physics (2007)
Ph.D. in Physics (2014)

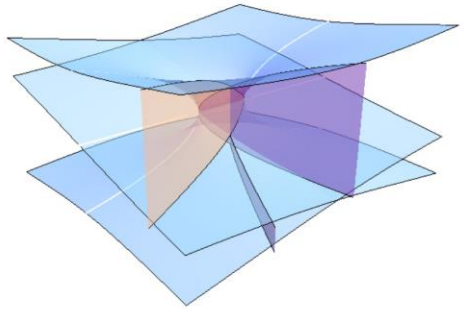
$$\Sigma_\rho := \left\{\lambda \,\middle|\, \det\left(\lambda\,\mathbb{I}_d - \rho(\varphi)\right) = 0\right\} \subset T^*C$$

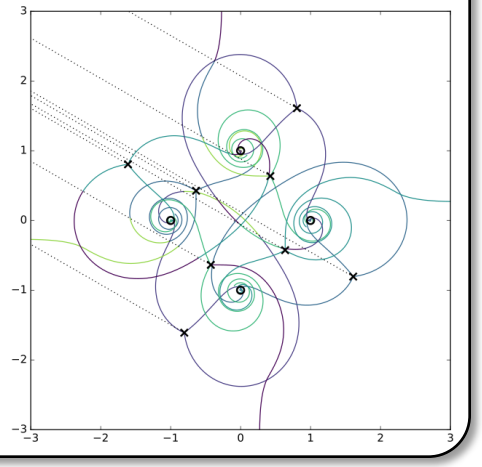
$$\pi^{-1}(z) = \left\{\lambda_z \in T_z^*C \,\middle|\, \prod_{j=1}^{d} \left(\lambda_z - x_j(z)\,\mathrm{d}z\right) = 0\right\}$$

$$x_j(z)\,\mathrm{d}z = \langle \nu_j, \varphi(z)\rangle \in \mathbb{C},$$

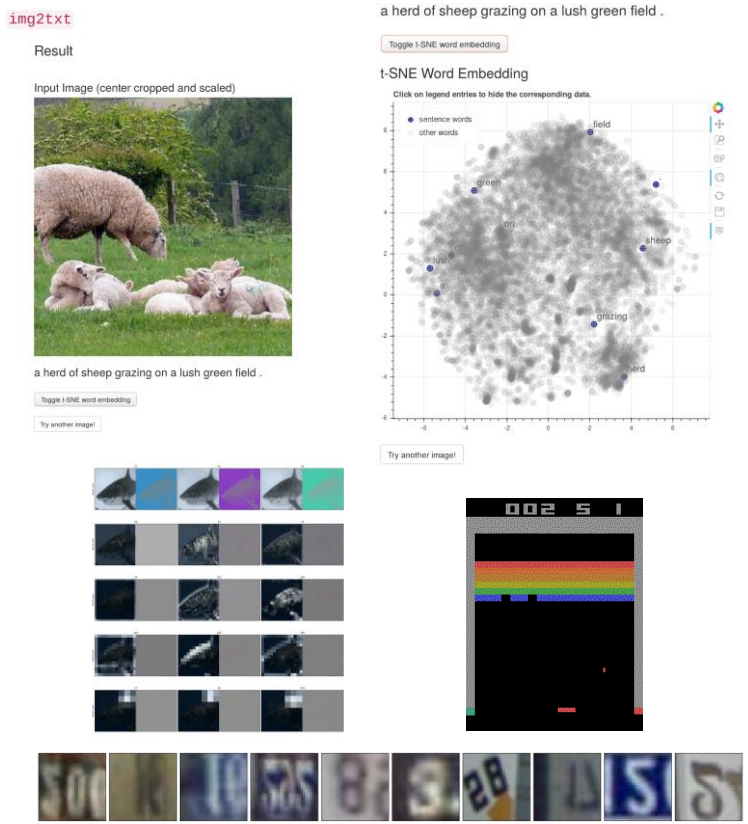
$$H_1(\Sigma_\rho, \mathbb{Z}) \,/\, \ker(Z) \simeq \hat{\Gamma}$$

Postdoc in high-energy theory (2014 ~ 2017)





Insight AI Fellow (2017)



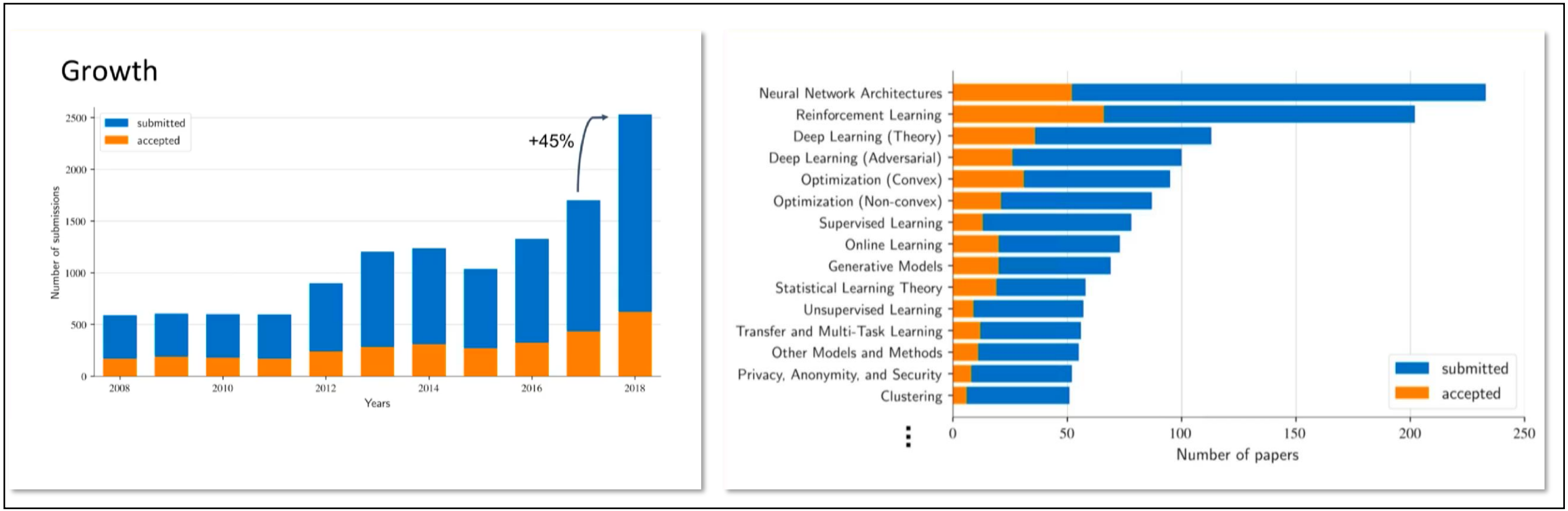
My research interest in machine learning

- Geometric deep learning
  - Hyperbolic embedding
  - Machine learning on graphs
- Replacing heuristics with learned models
  - Replacing heuristic algorithm and data structure with learned models
  - Identifying heuristics in computer architecture and hardware to replace with ML models.
- Understanding latent spaces of generative models
  - Disentangling and factorizing
  - Geometry of latent space manifold

ICML is one of the major ML conferences.

Francis Bach, Opening remarks, ICML 2018

Trends



Francis Bach, Opening remarks, ICML 2018

So many talks!

- Tuesday – 9 tutorials in 3 parallel sessions
- Wednesday to Friday – 10 parallel sessions, 200+ talks each day
- Friday to Sunday – 67 workshops in parallel
- No way to cover them all in this talk!
  - Other resources
    - David Abel's highlights (https://david-abel.github.io/blog/posts/misc/icml_2018.pdf)
      - Focus on RL
    - Full live streams (https://www.facebook.com/pg/icml.imls/videos)
    - All conference videos will be available after about a month.
  - I will cover highlighted topics and then focus on topics of my research interest.

Plan

- Highlighted topics
  - Security of ML
  - Fair ML
  - Bayesian Inference
  - Theory of Deep Learning
- Interesting topics
  - Geometry and Deep Learning
  - Replacing Heuristics with Machine Learning
  - Understanding Latent Spaces of Generative models
- Other topics

Plan

- **Highlighted topics**
  - Security of ML
  - Fair ML
  - Bayesian Inference
  - Theory of Deep Learning
- Interesting topics
  - Geometry and Deep Learning
  - Replacing Heuristics with Machine Learning
  - Understanding Latent Spaces of Generative models
- Other topics

Security of ML

- Keynote speech
  - *AI and Security: Lessons, Challenges and Future Directions*, Dawn Song
- Best paper award
  - *Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples*, Anish Athalye, Nicholas Carlini, David Wagner
- Debate

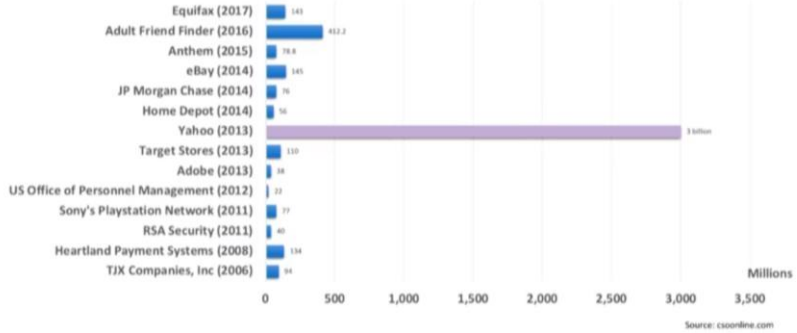
Keynote, D. Song

Massive DDoS Caused by IoT Devices (Mirai Botnet)

Geographical distribution of Mirai bots in recent DDoS attack

- Over 400,000 compromised IoT devices over 160 countries
  - Security cameras/webcams/baby monitors
  - Home routers
- One of the biggest DDoS attacks in history
  - Over 1Tbps combined attack traffic

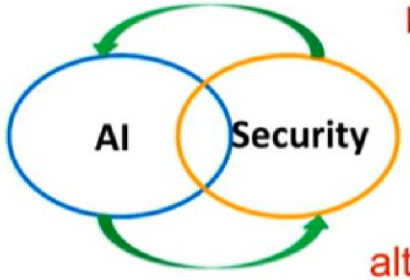
Biggest Data Breaches Of the 21st Century



Attacks Entering New Landscape

Ukraine power outage by cyber attack impacted over 250,000 customers

Millions of dollars lost in targeted attacks in SWIFT banking system



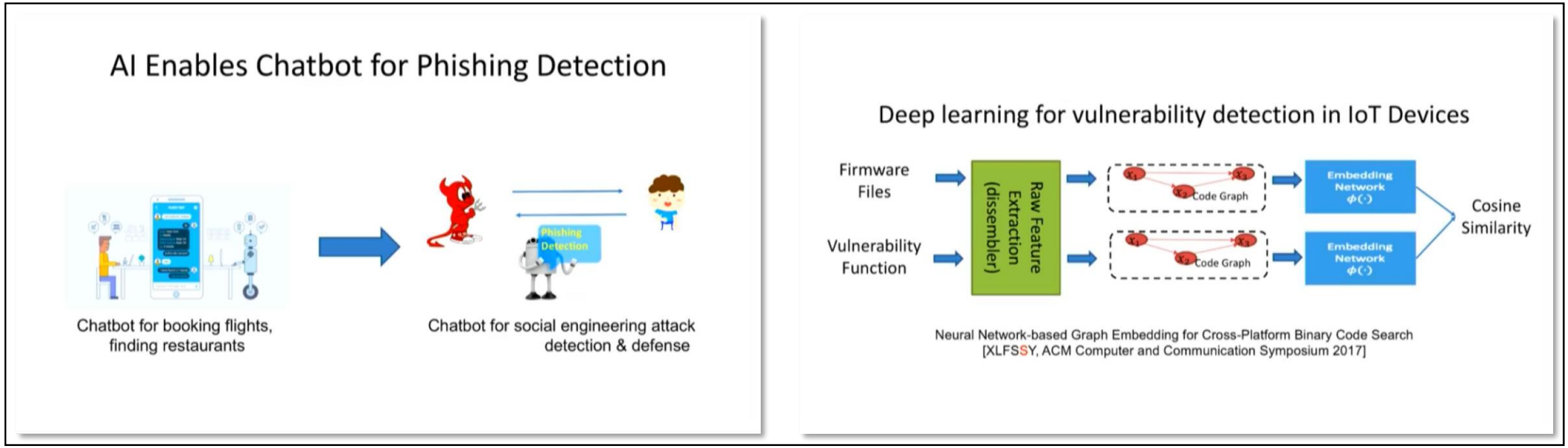
How will (in)security impact the deployment of AI?

How will the rise of AI alter the security landscape?

AI and Security: Lessons, Challenges and Future Directions, D. Song, ICML 2018

Keynote, D. Song

- Vulnerability detection using deep learning
  - Chatbot phishing detection
  - IoT device vulnerability detection



AI and Security: Lessons, Challenges and Future Directions, D. Song, ICML 2018

Keynote, D. Song

- Machine learning in the presence of attackers
  - Adversarial examples / data poisoning

AI and Security: Lessons, Challenges and Future Directions, D. Song, ICML 2018

  - Protecting Privacy ("Data is the new oil.")
    - Does neural network remember the training data?
    - If yes, can an attacker extract it by querying the model?
  - No sufficient defense today

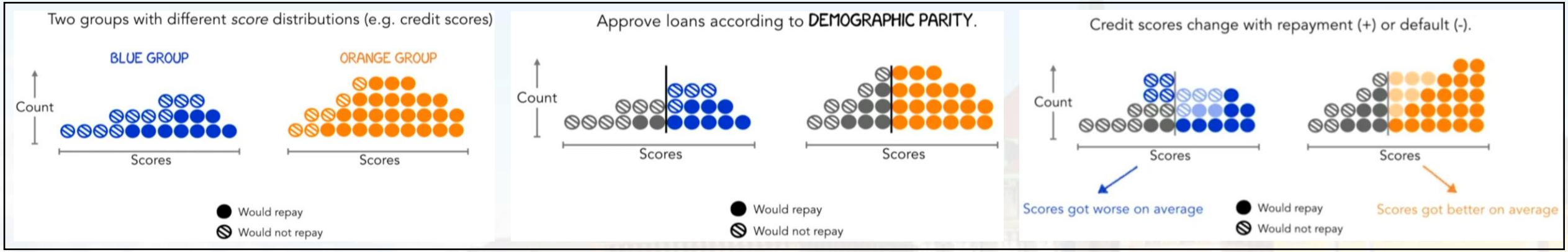
Best paper award, N. Carlini

- *Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples*, Anish Athalye, Nicholas Carlini, David Wagner
- Among 13 defense papers at ICLR 2018,
  - 9 are white-box, non-certified.
  - 6 of them are broken, 1 of them is partially broken.
- The threat model must assume the attacker has read the paper and knows the defender is using those techniques to defend.
- Learn to break defenses before trying to build them.
  - If you cannot break the state-of-the-art, you are unlikely to be able to build on it.

Debate

- Proposition: "The vulnerabilities of present machine learning systems are so critical that we should not allow their general deployment in real-world settings.“
  - Aleksander Madry (Affirmative)
  - Alhussein Fawzi (Affirmative)
  - Percy Liang (Negative)
  - Aditi Raghunathan (Negative)
- Whether to put resources in advancing the ML frontier vs. in securing the ML tech as early as possible.

Fair ML



Delayed Impact of Fair Machine Learning, Lydia Liu et al., ICML 2018

- Best paper award
  - *Delayed Impact of Fair Machine Learning*, Lydia Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt
    - Fairness does not help protected groups and may actually harm them.
- Workshop
  - Fairness, Accountability, and Transparency in Machine Learning

Bayesian Inference

- Tutorial
  - *Variational Bayes and Beyond: Bayesian Inference for Big Data*, T. Broderick
- Keynote
  - *Intelligence per Kilowatthour*, M. Welling.
- (Conditional) Neural Process

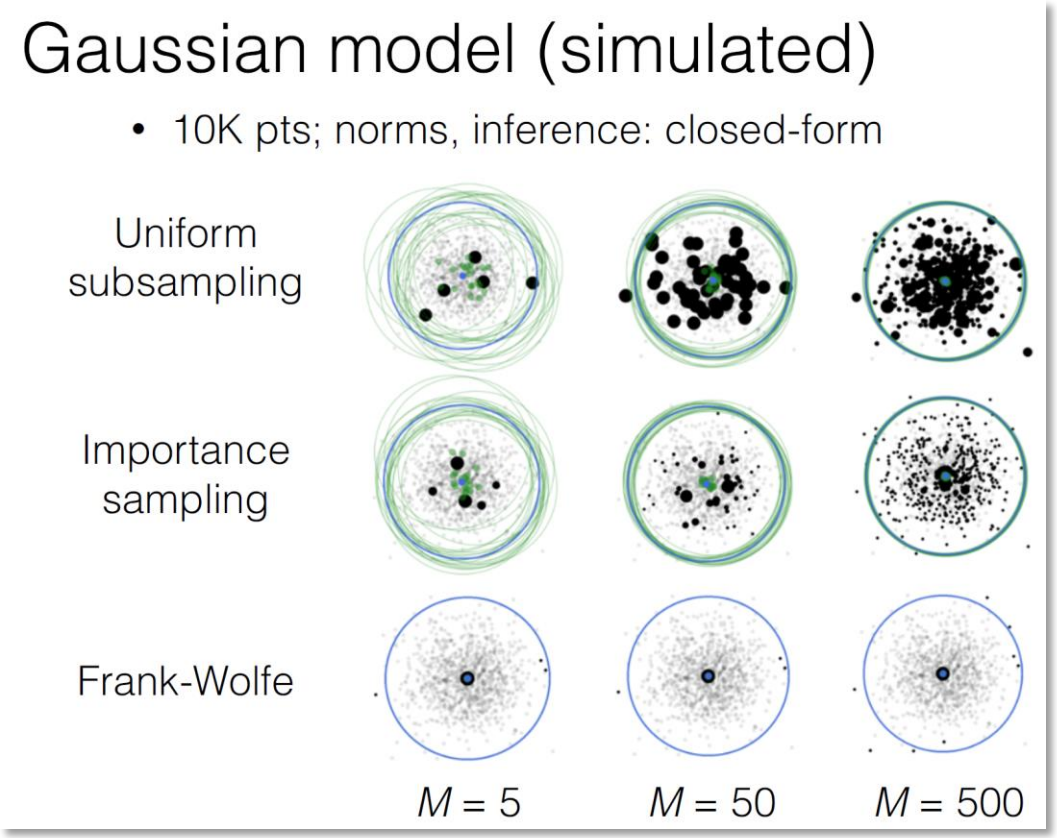
Tutorial, T. Broderick

- Bayesian inference still going strong these days.

Variational Bayes and Beyond: Bayesian Inference for Big Data, T. Broderick, ICML 2018

Tutorial, T. Broderick

- Slow for modern problems: large data, large dimensions
- Approximate Bayesian inference
  - Variational Bayesian inference
  - Mean-field variational Bayes
- Preprocess data to build Bayesian coreset.
  - a smaller, weighted dataset



Variational Bayes and Beyond: Bayesian Inference for Big Data, T. Broderick, ICML 2018

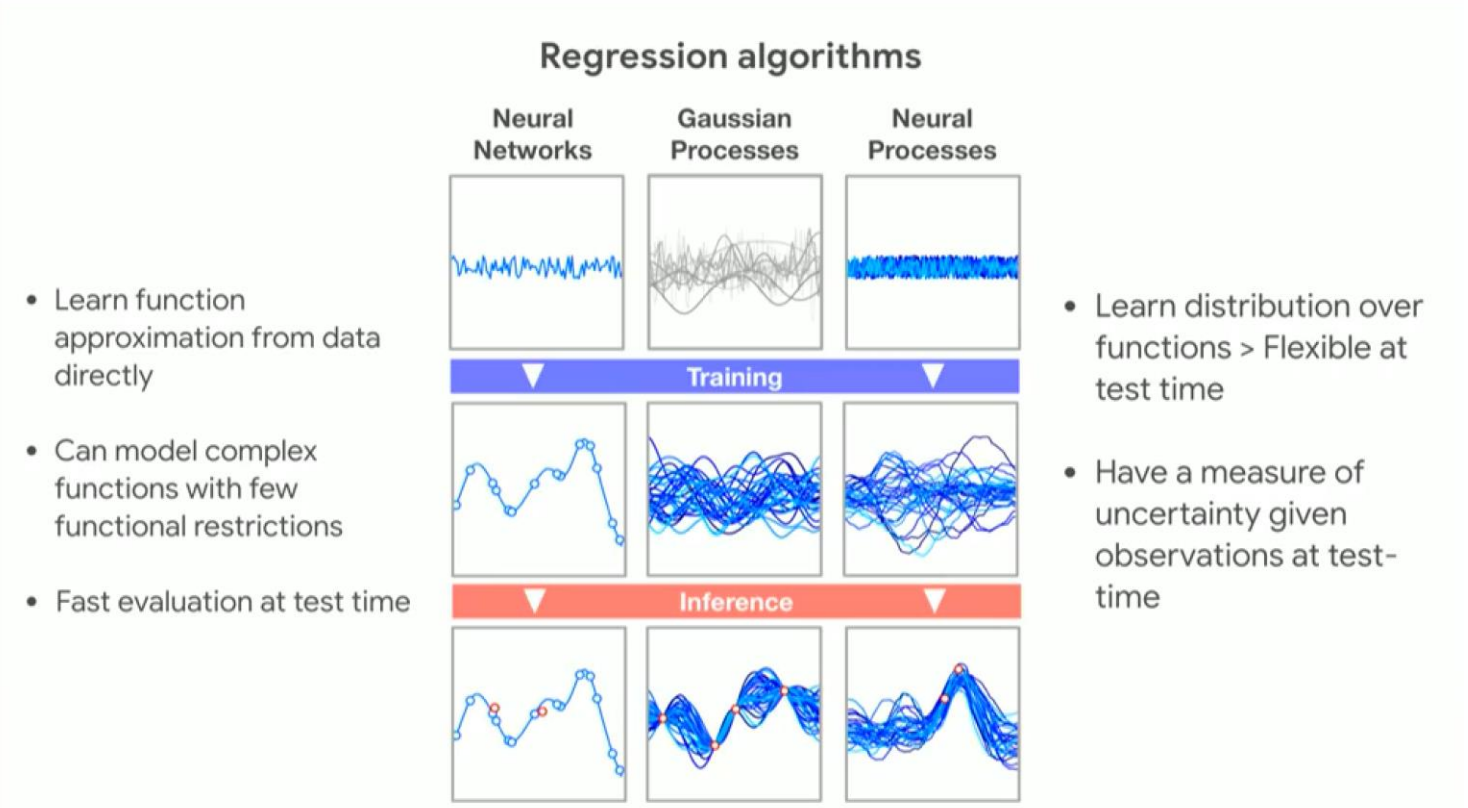
Keynote, M. Welling

Intelligence per Kilowatthour, M. Welling, ICML 2018

- "Value created by AI must exceed the cost to run the service."
- For energy efficient deep learning, use Bayesian deep learning for model compression.
- Use posterior distribution of parameters after data.
  - Variational dropout posterior results in sparse models.

Neural process

- *Conditional Neural Processes*, Marta Garnelo et al.
- Learn distribution of functions.
- Trade-off between Gaussian process and conditional neural process



#	Random Context			Ordered Context		
	10	100	1000	10	100	1000
kNN	0.215	0.052	0.007	0.370	0.273	0.007
GP	0.247	0.137	**0.001**	0.257	0.220	**0.002**
CNP	**0.039**	**0.016**	0.009	**0.057**	**0.047**	0.021

Table 1. Pixel-wise mean squared error for all of the pixels in the image completion task on the CelebA data set with increasing number of context points (10, 100, 1000). The context points are chosen either at random or ordered from the top-left corner to the bottom-right. With fewer context points CNPs outperform kNNs and GPs. In addition CNPs perform well regardless of the order of the context points, whereas GPs and kNNs perform worse when the context is ordered.

Theory of Deep Learning

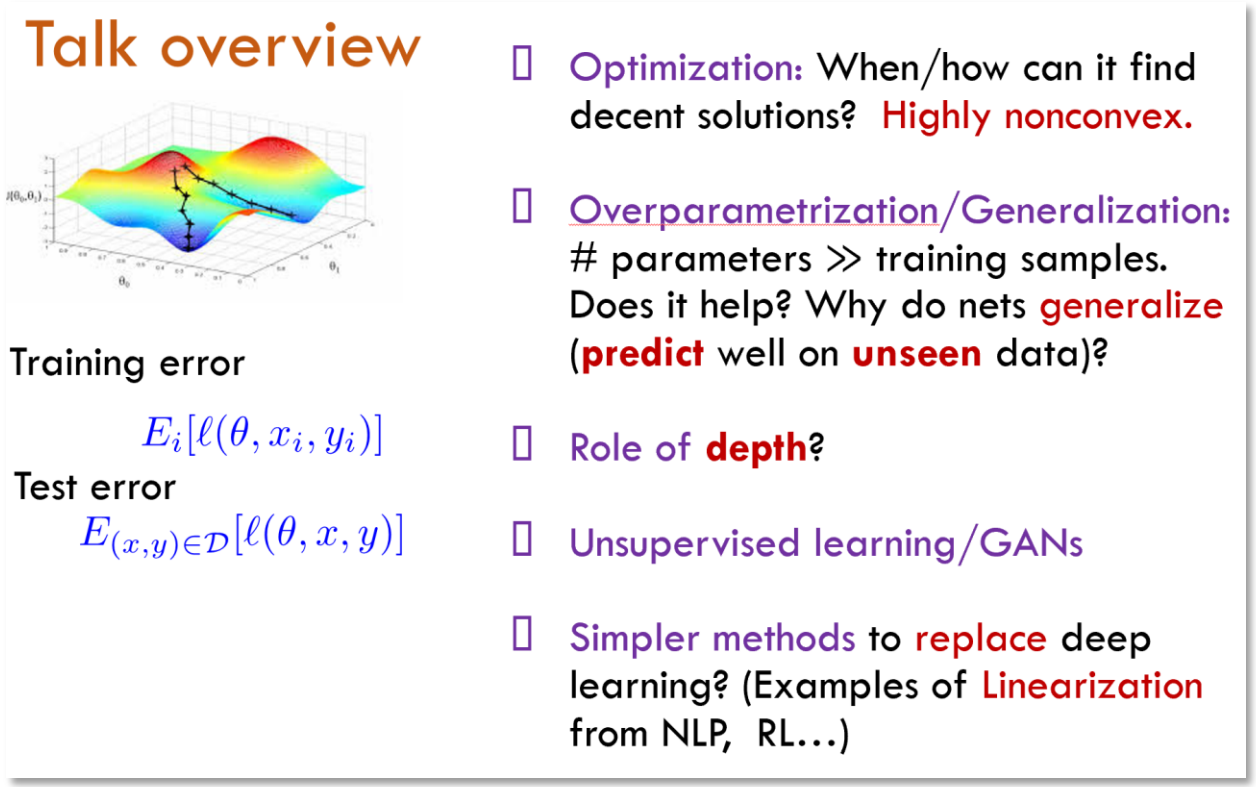
- Conference
  - *Understanding the Loss Surface of Neural Networks for Binary Classification*, Shiyu Liang, Ruoyu Sun, Yixuan Li, Rayadurgam Srikant.
  - *Tropical Geometry of Deep Neural Networks*, Liwen Zhang, Gregory Naitzat, Lek-Heng Lim.
- Tutorial
  - Toward Theoretical Understanding of Deep Learning, S. Arora.

Mathematical models

- *Understanding the Loss Surface of Neural Networks for Binary Classification*, Shiyu Liang, Ruoyu Sun, Yixuan Li, Rayadurgam Srikant.
  - Focus on the training of neural networks for binary classification.
  - Provide conditions under which the training error is zero at all local minima.
    - the activation function have to be increasing and strictly convex,
    - the neural network should either be single-layered or is multi-layered with a shortcut-like connection,
    - and the loss function should be a smooth version of hinge loss.
- *Tropical Geometry of Deep Neural Networks*, Liwen Zhang, Gregory Naitzat, Lek-Heng Lim.
  - Identify a connection between a class of neural networks and tropical algebraic geometry.
  - The family of feedforward neural networks with ReLU activation are equivalent to the family of tropical rational maps.

Tutorial, S. Arora.

- Optimization
  - 2nd order optimization does not find better quality model.
- Overparameterization
  - No meaningful estimation of the capacity of deep models.
- Generalization
  - Hard to make quantitative statement about flat minima.
- Role of depth
  - Currently not within reach of theory.
- GAN
  - Efforts to understand mode collapse problem theoretically.
- Most of the studies are post-mortem.



Toward Theoretical Understanding of Deep Learning, S. Arora, ICML 2018

Tutorial, S. Arora.

Food for thought…

Maximizing log likelihood (presumably approximately) may lead to little usable insight into the data.

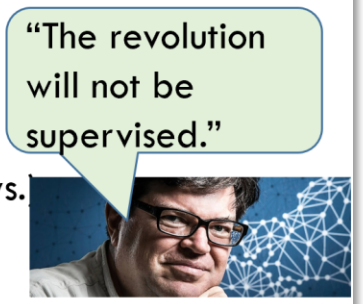
How to define utility of GANs (if not as distribution learners)?

Need to define unsupervised learning using a "utility" approach (What downstream tasks are we interested in and what info do they need about X?)

(Similar musings on INFERENCE blog, April'18.
e.g., What would a "representation learning competition" look like?)

What to work on (suggestions for theorists)

1. Use Physics/PDE insights, such as calculus of variations (Lagrangians, Hamiltonians, etc.)
2. Look at unsupervised learning (Yes, everything is NP-hard and new but that's how theory grows.
3. Theory for Deep Reinforcement learning. (Currently very little.)
4. Going beyond 3), design interesting models for interactive learning (of language, skills, etc.). Both theory and applied work here seems to be missing some basic idea. (Theory focuses on simple settings like linear classifiers/clustering.)



Toward Theoretical Understanding of Deep Learning, S. Arora, ICML 2018

Physicists @ ICML

- Physicists in the wild grazing on the grass… or just having a lunch together.

Plan

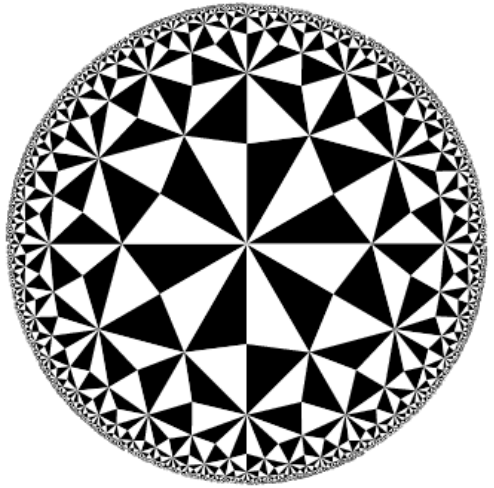
- Highlighted topics
  - Security of ML
  - Fair ML
  - Bayesian Inference
  - Theory of Deep Learning
- **Interesting topics**
  - Geometry and Deep Learning
  - Replacing Heuristics with Machine Learning
  - Understanding Latent Spaces of Generative models
- Other topics

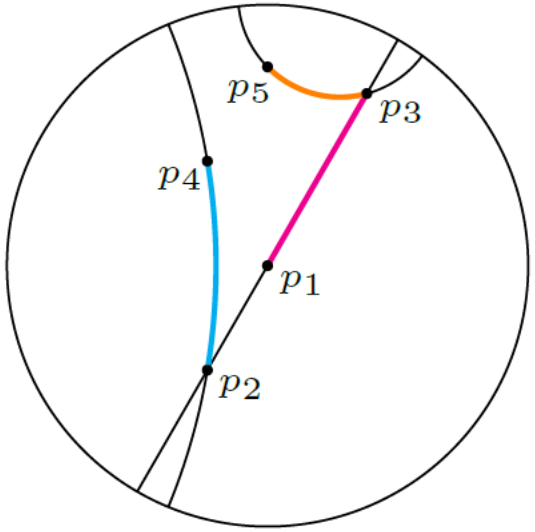
Geometry and Deep Learning

- Geometric deep learning
  - *"An umbrella term for emerging techniques attempting to generalize (structured) deep neural models to non-Euclidean domains such as **graphs** and **manifold**."*
    - From *Geometric deep learning: going beyond Euclidean data*, M. Bronstein et al., IEEE Signal Processing Magazine (Volume: 34, Issue: 4, July 2017)
    - Examples
      - Social networks
      - Sensor networks
      - Functional networks in brain imaging
      - Regulatory networks in genetics
      - Meshed surfaces in computer graphics.
  - Hyperbolic embedding
    - Learning continuous hierarchy
    - Hyperbolic entailment cone
- Equivariance in deep learning

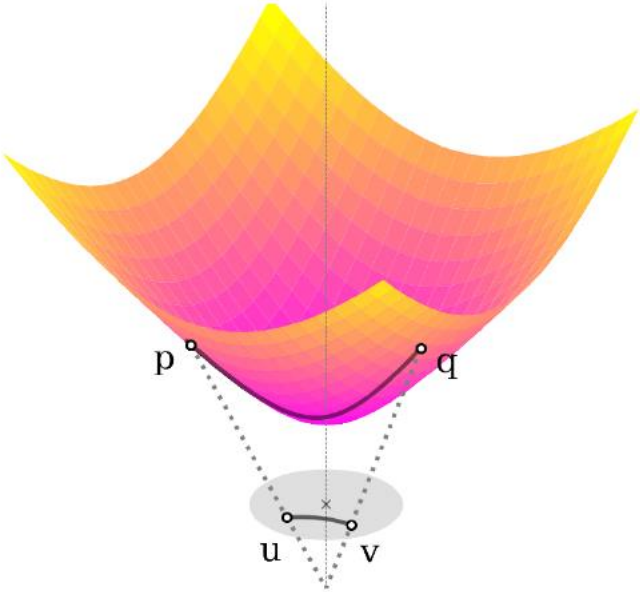
Hyperbolic embedding

- *Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry*, Maximillian Nickel, Douwe Kiela.
  - Poincare model for visualization
  - Lorentz model for Riemannian optimization

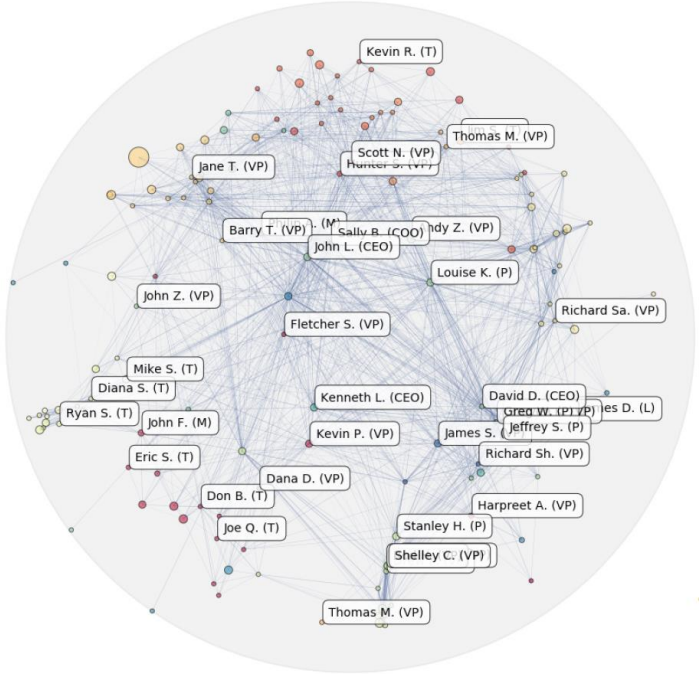




(a) Geodesics in the Poincaré disk.

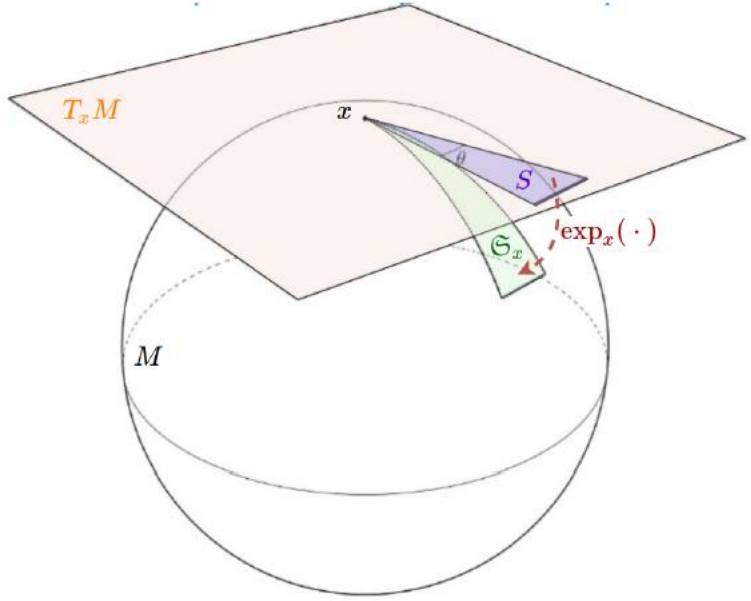


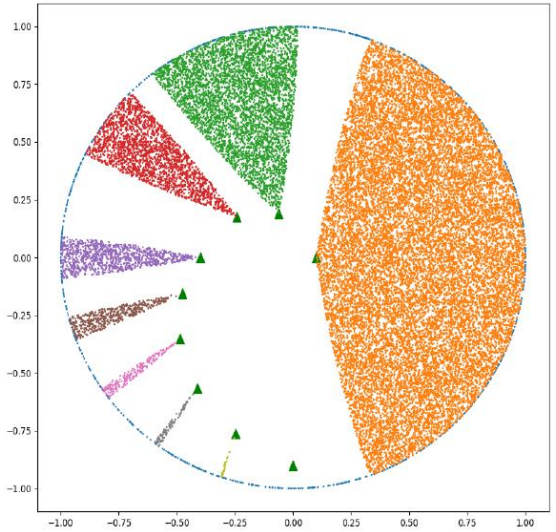
(b) Lorentz model of hyperbolic geometry.

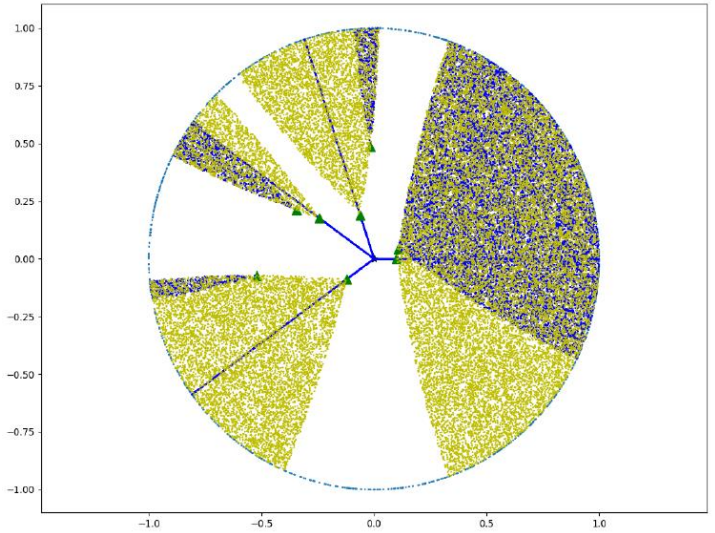


Hyperbolic embedding

- *Hyperbolic Entailment Cones for Learning Hierarchical Embeddings*, Octavian-Eugen Ganea, Gary Becigneul, Thomas Hofmann.
- Embed directed acyclic graphs (DAGs) using nested Riemannian convex cones.
  - Mathematically defined entailment cones in a closed form.







Deep learning on graphs

- *Stochastic Training of Graph Convolutional Networks with Variance Reduction*, Jianfei Chen et al.
  - Problem of GCN: large receptive field, recursive neighbors.
  - Use history rather than sampling.
- *Representation Learning on Graphs with Jumping Knowledge Networks*, Keyulu Xu et al.
  - For GCN using neighborhood aggregation (or message passing), 2-layer is best, the deeper the worse.
  - Theorem: k-layer GCN is equivalent to k-step random walk.
  - Suggests JK-net based on the observation.

Deep learning on graphs

- *NetGAN: Generating Graphs via Random Walks*, A. Bojchevski, O. Shchur et al.
  - Generative model for graphs that captures properties of real-world graph.
  - Learn implicit model of random walk distribution on graphs using GAN.
  - Random walk is a Markov process and therefore has no memory. Why more than 2 steps? For Generalization!
- *GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models*, Jiaxuan You et al.
  - Generating realistic graph by modeling graphs as sequences and learning the probability distribution of graphs.

Equivariance in DL models

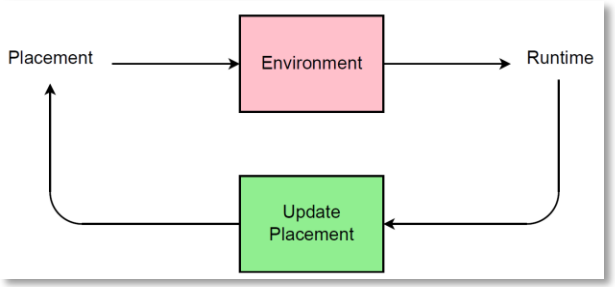
- *Towards learning with limited labels: Equivariance, Invariance, and Beyond*, Workshop @ ICML 2018
  - Equivariance : $\Phi(T_g x) = T'_g \Phi(x)$ for $g \in G$, where x is an input, Φ is a non-linear function represented by a deep network, and G is the symmetry group of the input space.
    - Leads to the consideration of fiber bundle (G-bundle).
  - Imposing invariance too soon results in loosing information and requiring learning many kernels that are transformation of each other.
  - *The General Theory of Equivariant Convolutional Networks*, Taco Cohen.
  - *Capsule networks and transformation extrapolation for learning from limited data*, Nicholas Frosst.

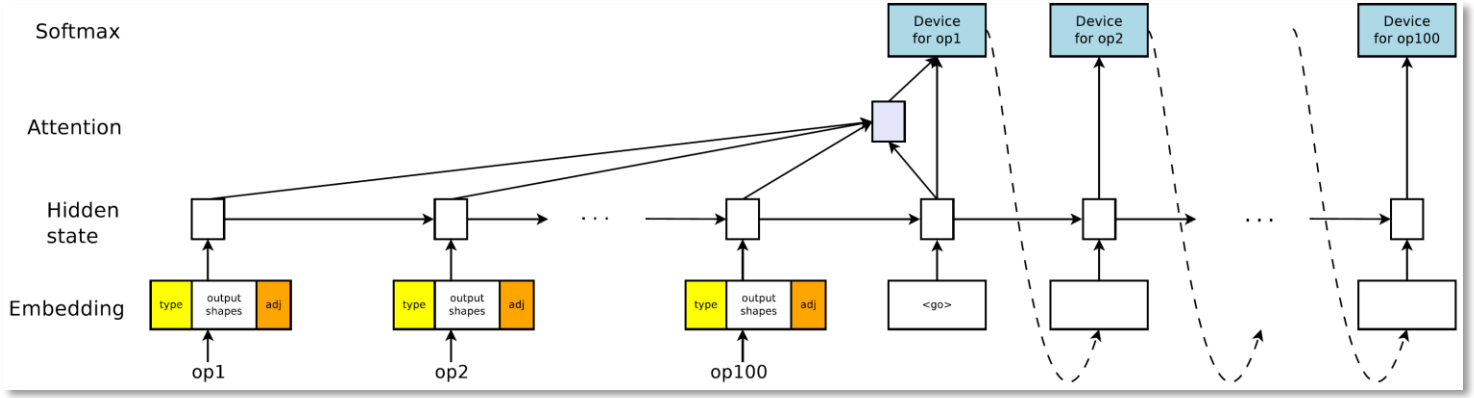
Replacing Heuristics with ML SW & HW

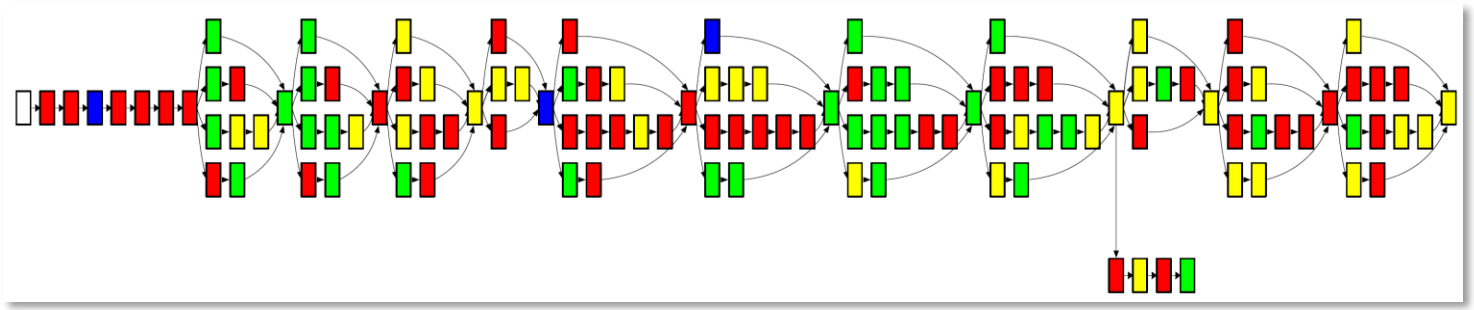
- Replacing heuristic algorithm and data structure with learned models
- Domain specific architecture
- RL-based memory controller
- Learning memory access pattern
- Imitation learning

Replacing heuristic algorithm with ML models

- *Device Placement Optimization with Reinforcement Learning*, Azalia Mirhoseini and Hieu Pham et al., ICML 2017
  - Learn to optimize device placement for TF computation graphs.
  - RL-based placement of Inception-v3 achieves the improvement of 19.7% in running time compared to expert-designed placement.

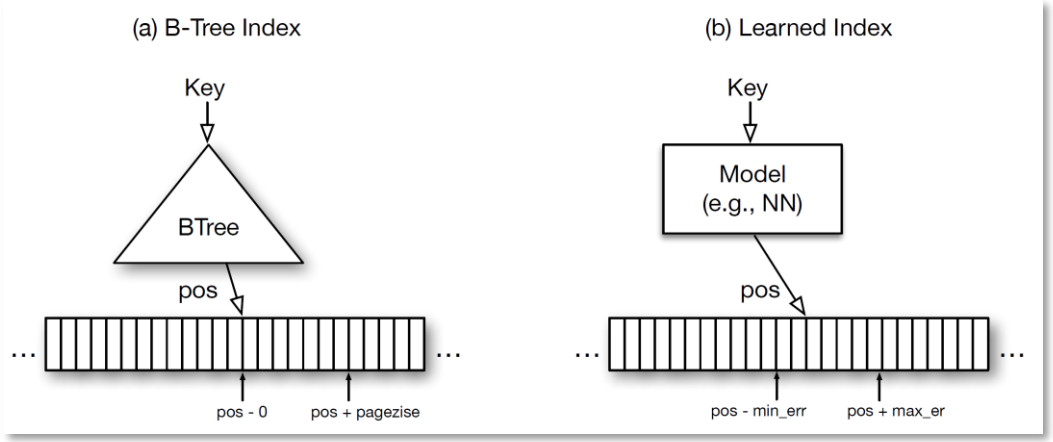






Replacing heuristic data structure with ML models

- *The Case for Learned Index Structures*, T. Kraska et al., arXiv:1712.01208
  - A B-Tree-Index can be seen as a model to map a key to the position of a record within a sorted array.
  - Index structures can be replaced with deep-learning models.
  - Learned index structures using neural nets outperform cache-optimized B-Trees by up to 70% in speed while saving an order-of-magnitude in memory over several real-world data sets.



Fireside chat with Jeff Dean at Google

- *Systems and Machine Learning Symbiosis*, Jeff Dean, June 26 @ Google Campus Seoul
  - Q1: In addition to replacing algorithmic index data structures like B-Tree with machine learning models. I wonder if there has been additional research efforts regarding replacing other data structures based on heuristics with learned models.
  - A1: There are many heuristics in modern computer systems, for example cache policies, and replacing these with ML/RL will be interesting.
  - Q2: TPU is a great example of an ASIC specialized for the acceleration of training and inference of DL models. What other ASICs can we expect to replace and accelerate current algorithm-based data structures with machine learning models?
  - A2: It is important for the development of an ASIC to hit the sweet spot of specialization and wide applicability, and for now accelerating linear algebra operations seems to be the most appropriate one, albeit it may take a form of low-power edge device instead of TPUs at datacenters.

Domain specific architecture

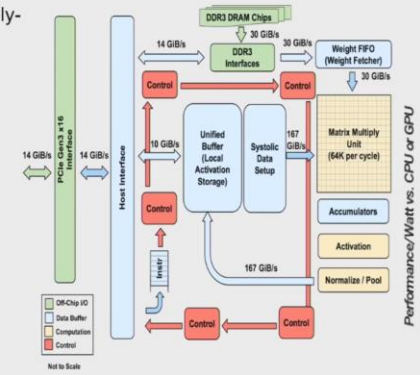
- John Hennessy, ACM A.M. Turing Award Lecture @ ISCA 2018
  - Performance improvements are at a standstill.
  - Lots of opportunities, but new approach to computer architecture is needed.

Domain Specific Architectures (DSAs)

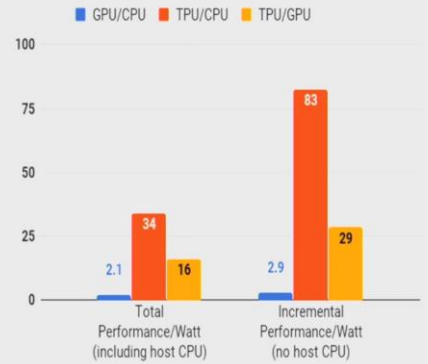
- Achieve higher efficiency by tailoring the architecture to characteristics of the domain
  - Not one application, but a domain of applications
    -Different from strict ASIC
  - Requires more domain-specific knowledge then general purpose processors need
- Examples:
  - Neural network processors for machine learning
  - GPUs for graphics, virtual reality
  - Programmable network switches and interfaces

TPU: High-level Chip Architecture

- The Matrix Unit: 65,536 (256x256) 8-bit multiply-accumulate units
- 700 MHz clock rate
- Peak: 92T operations/second
  - 65,536 * 2 * 700M
- >25X as many MACs vs GPU
- >100X as many MACs vs CPU
- 4 MiB of on-chip Accumulator memory
- 24 MiB of on-chip Unified Buffer (activation memory)
- 3.5X as much on-chip memory vs GPU
- Two 2133MHz DDR3 DRAM channels
- 8 GiB of off-chip weight DRAM memory



Perf/Watt TPU vs CPU & GPU



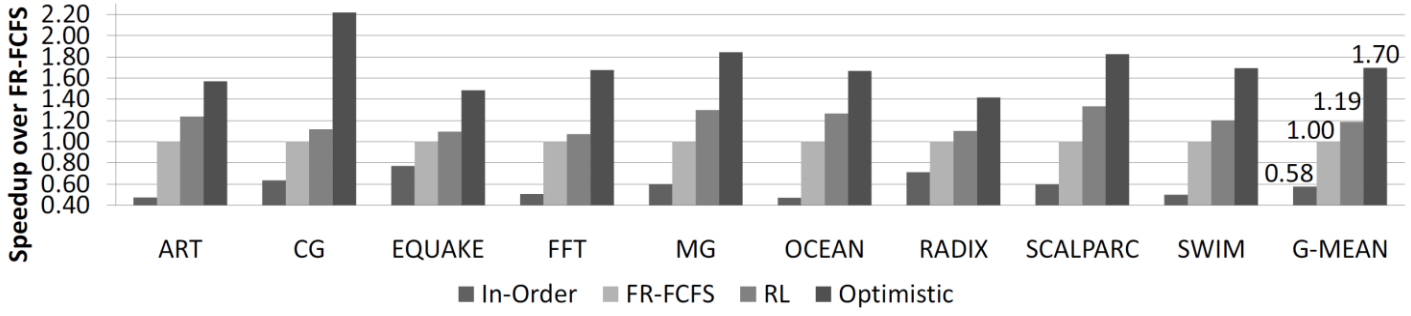
Measure performance of Machine Learning?

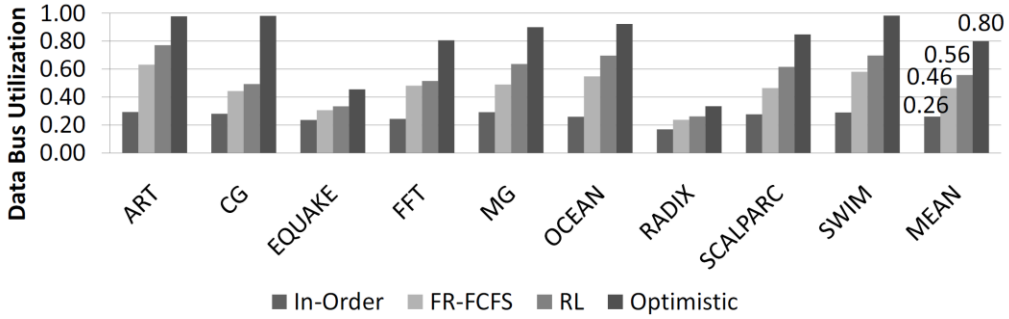
See MLPerf.org ("SPEC for ML")

- Benchmark suite being developed by
  - ≥7 companies and ≥5 universities
  - To be released 7/1/18

RL-based memory controller

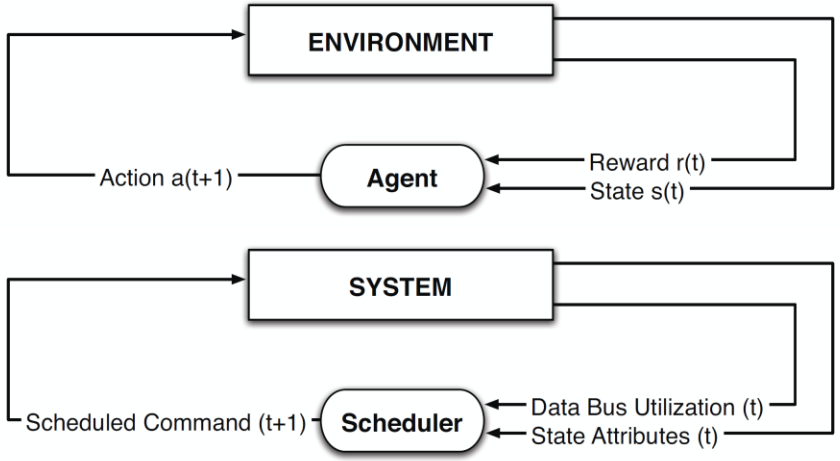
- *Self-Optimizing Memory Controllers: A Reinforcement Learning Approach*, E. Ipek et al., ISCA 2008.
  - Propose to design the memory controller as an RL agent to learn an optimal memory scheduling policy for chip multiprocessors (CMP).
  - Compared to fixed access scheduling policies (i.e. FR-FCFS), RL-based controller improves
    - the performance of a set of parallel applications by 19% on average (up to 33%),
    - and DRAM bandwidth utilization by 22% on average.

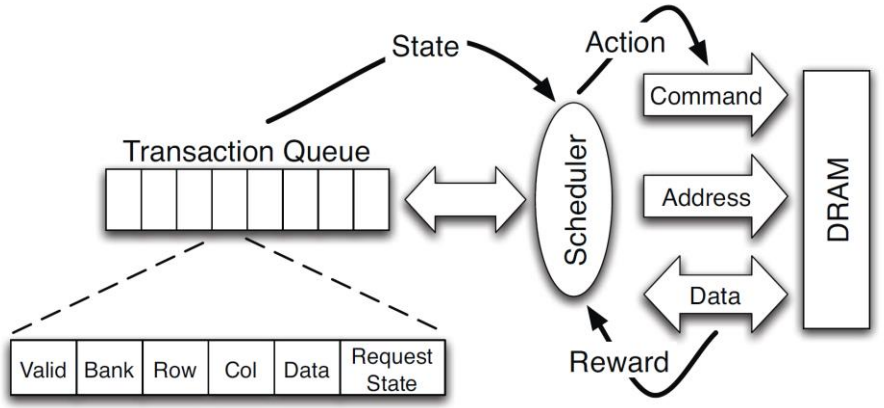




RL-based memory controller

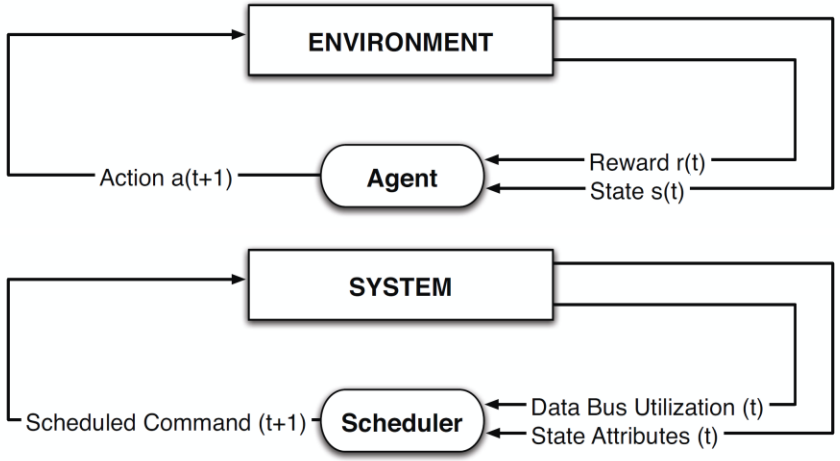
- Use Q-learning to train an RL-based memory controller.
- Allow HW designer to focus on
  - **what** performance target to accomplish
  - and **what** system variables might be useful
  - rather than devising a fixed policy that describes **exactly how** the controller should accomplish the target.

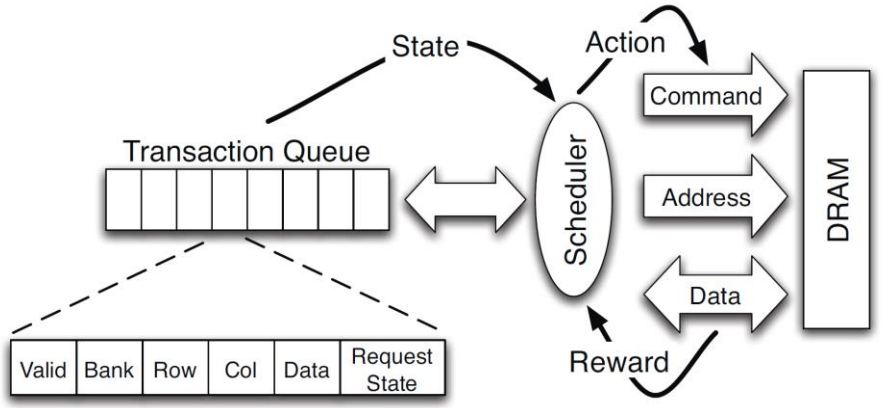




RL-based memory controller

- How self-optimizing memory controller fits into RL framework:
  - State: six attributes selected via **feature engineering**.
    - # of reads, writes, load misses in the transaction queue, the criticality of each request, whether a given request would hit in the row buffer if serviced next, total # of reads and writes pending for each row and bank.
  - Action: legal DRAM commands – precharge, activate, read, write.
  - Reward: utilization of the data bus.
    - 1 for issuing a command of read or write, 0 for others.





DL-based memory prefetcher

- *Learning Memory Access Patterns*, Milad Hashemi et al.
  - Relate prefetching strategies to N-gram models in NLP.
  - Use RNNs to replace prefetchers based on prediction and heuristics.
  - Effectiveness of sequential learning algorithms in microarchitecture designs is still an open problem.
    - One of the authors says he got more questions than answers from this work.
  - Only simulation results, does not evaluate the HW design of the models.
    - Evaluates the precision and recall of cache hits/misses.
    - Unclear if DNNs can meet the latency demands.
  - Train-offline test-online model.
  - Shift the problem of prefetching from a memory capacity problem to a compute problem.

DL-based memory prefetcher

- Two input features are:
  - the sequence of cache miss addresses,
  - and the sequence of instruction addresses (program counters or PCs)
    - Can inform the model of patterns in the control flow.
- Address space extremely sparse, predict deltas instead.
  - $\Delta_N = \text{Addr}_{N+1} - \text{Addr}_N$
  - # of deltas is often orders of magnitude smaller.
  - Also works with address space layout randomization (ASLR)

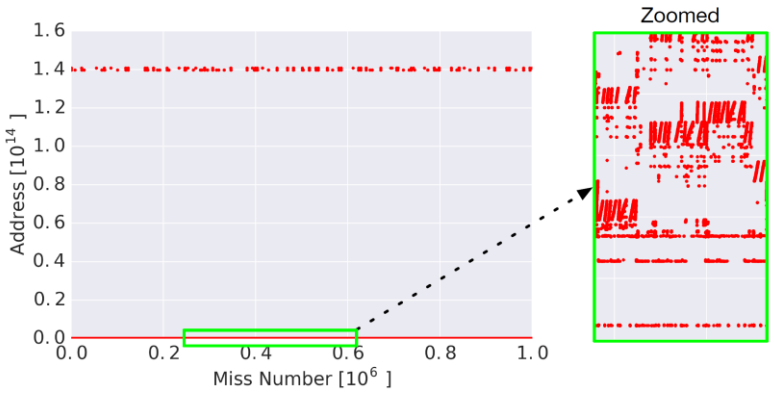
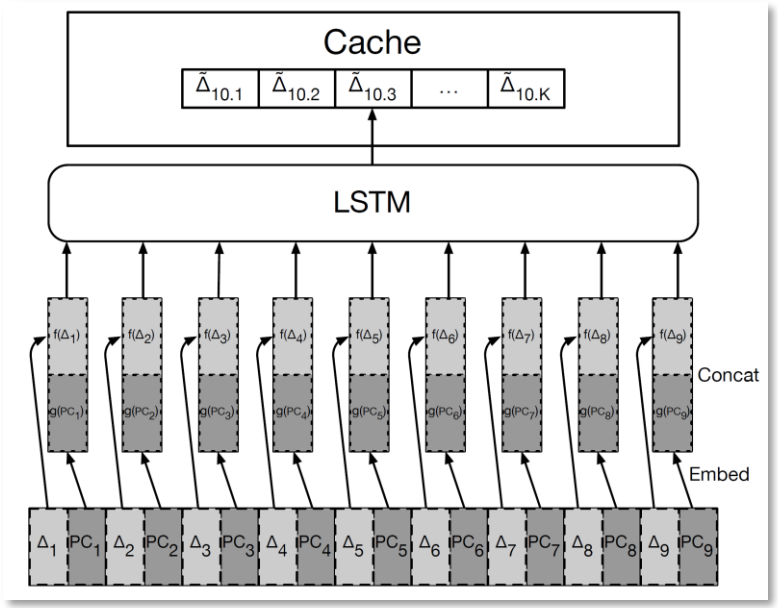


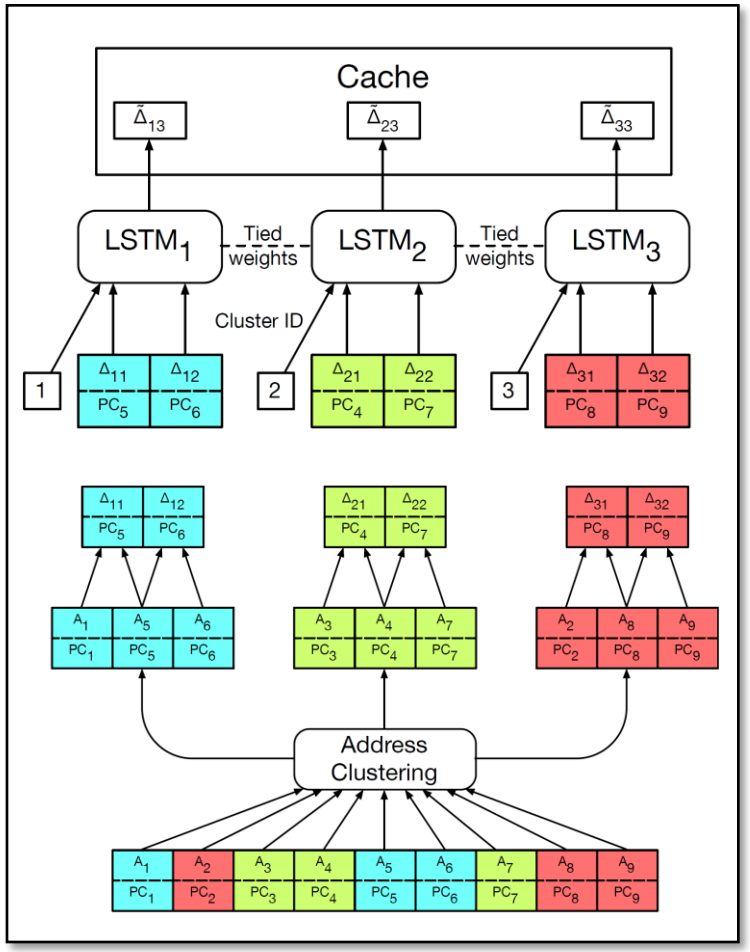
Table 1. Program trace dataset statistics. M stands for million.

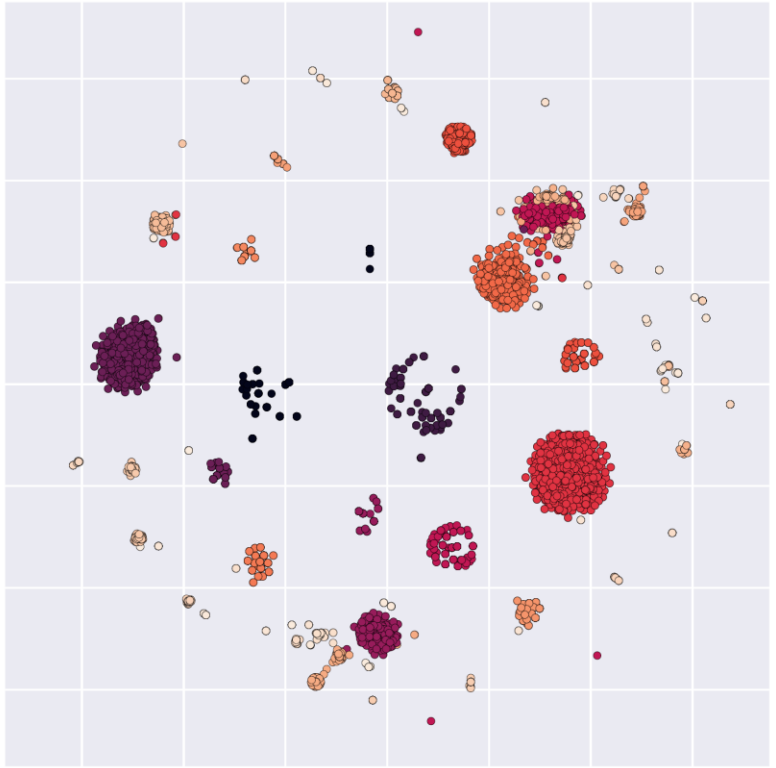
Dataset	# Misses	# PC	# Addrs	# Deltas	# Addrs 50% mass	# Deltas 50% mass
gems	500M	3278	13.11M	2.47M	4.28M	18
astar	500M	211	0.53M	1.77M	0.06M	15
bwaves	491M	893	14.20M	3.67M	3.03M	2
lbm	500M	55	6.60M	709	3.06M	9
leslie3d	500M	2554	1.23M	0.03M	0.23M	15
libquantum	470M	46	0.52M	30	0.26M	1
mcf	500M	174	27.41M	30.82M	0.07M	0.09M
milc	500M	898	3.74M	9.68M	0.87M	46
omnetpp	449M	976	0.71M	5.01M	0.12M	4613
soplex	500M	1218	3.49M	5.27M	1.04M	10
sphinx	283M	693	0.21M	0.37M	0.03M	3
websearch	500M	54600	77.76M	96.41M	0.33M	5186

DL-based memory prefetcher

- Two models: embedding LSTM and clustering + LSTM

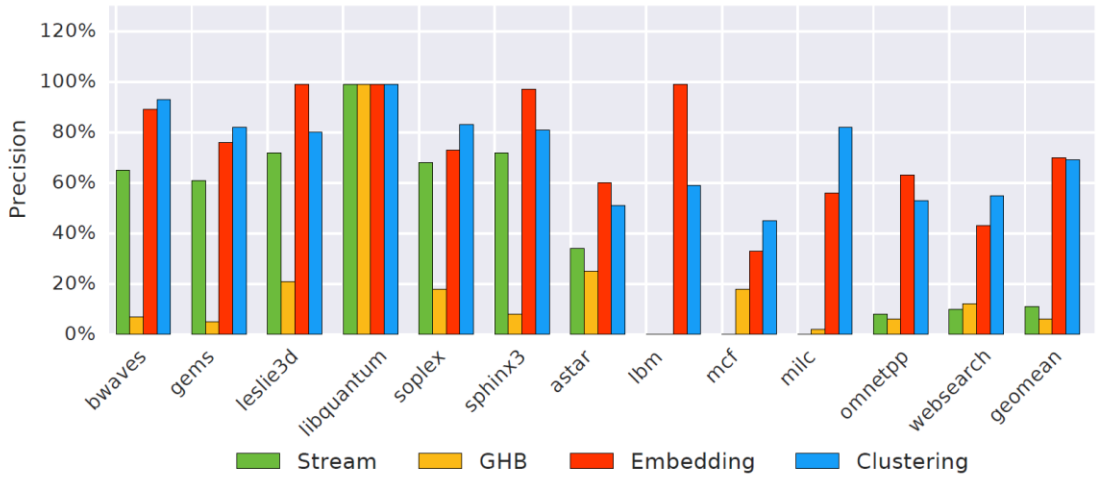


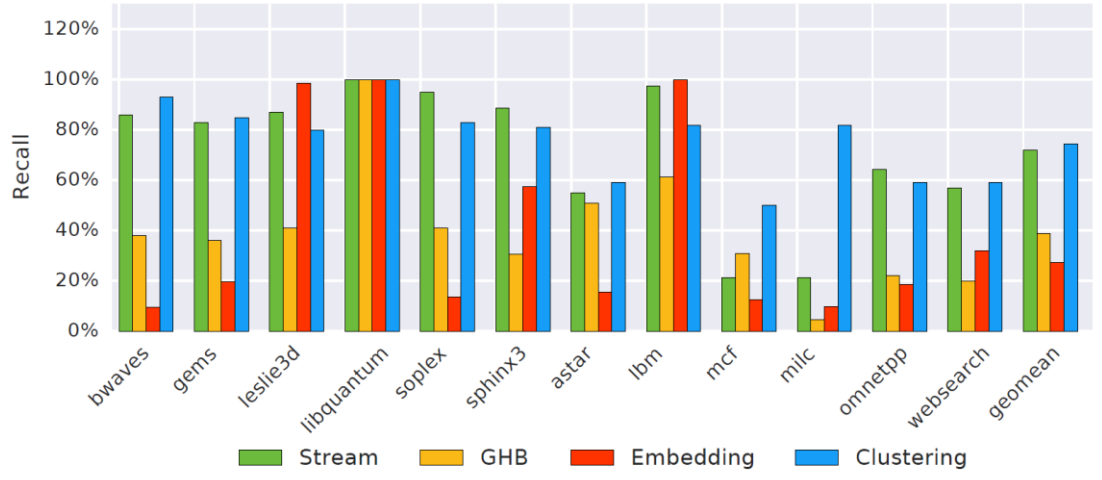




DL-based memory prefetcher

- The stream prefetcher achieves a high recall.
- LSTM models dominates in terms of precision.
  - Especially on Google's *websearch* workload dataset, added in addition to SPEC CPU2006 datasets.





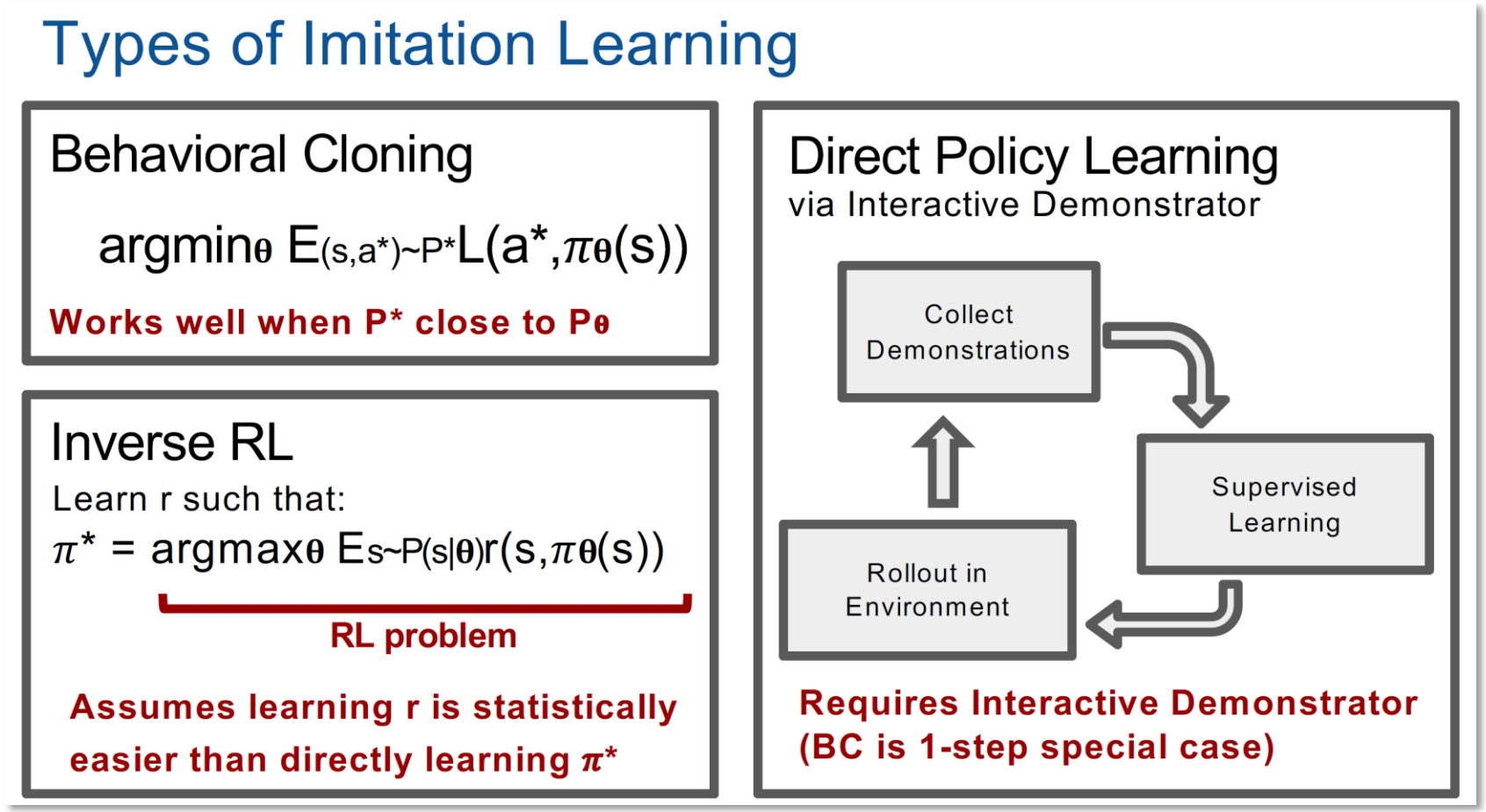
Other ML in microarchitecture

- Dynamic branch prediction with perceptrons, D. Jimenez and C. Lin, HPCA, 2001.
- Long short term based memory hardware prefetcher, Yuan Zeng, MEMSYS 2017.
- Reinforcement Learning-Assisted Garbage Collection to Mitigate Long-Tail Latency in SSD, Wonkyung Kang, Dongkun Shin, and Sungjoo Yoo, ACM 2017.

Imitation learning

- Is a powerful and practical alternative to reinforcement learning for learning sequential decision-making policies.
  - Reinforcement learning – learning policy without expert.
- Also known as learning from demonstrations or apprenticeship learning.
- What if we consider a heuristic as a suboptimal expert and train a policy that outperforms the expert?

Imitation learning



Imitation Learning, Y. Yue and H. M. Le, ICML 2018

Imitation learning

- Behavioral cloning
  - Supervised learning
  - Limitation: expert makes no mistake!
- Direct policy learning via interactive expert
  - Generalization of BC
  - Can query expert any time
  - Data aggregation and policy aggregation

Imitation learning

Direct Policy Learning via Interactive Expert

Reduction to sequence of supervised learning problems

- Constructed from roll-outs of previous policies
- Requires interactive expert feedback

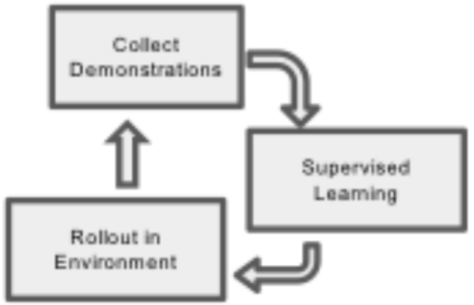
Two approaches: Data Aggregation & Policy Aggregation

- Ensures convergence
- Motivated by different theory

Not covered:

- What is expert feedback & loss function?

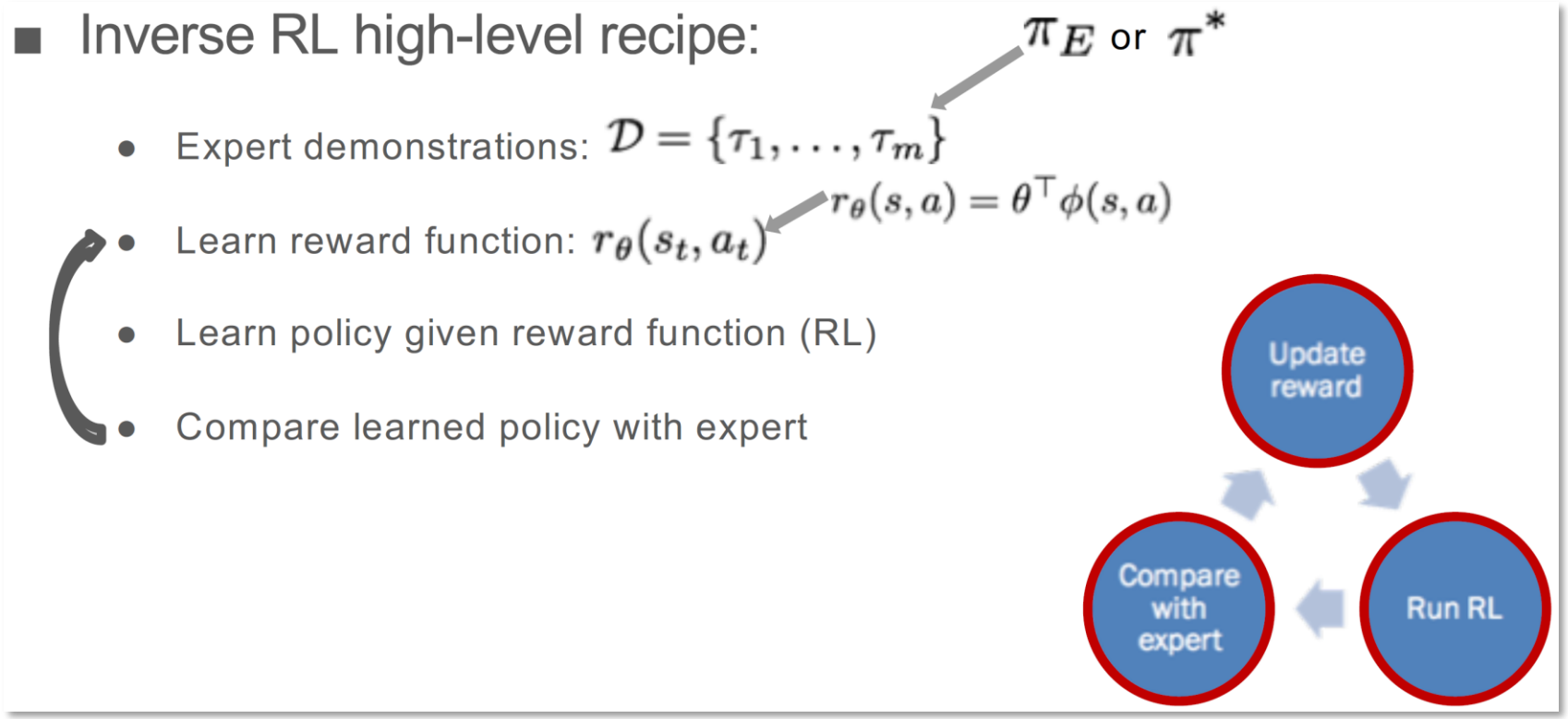
Depends on application



Imitation Learning, Y. Yue and H. M. Le, ICML 2018

Imitation learning

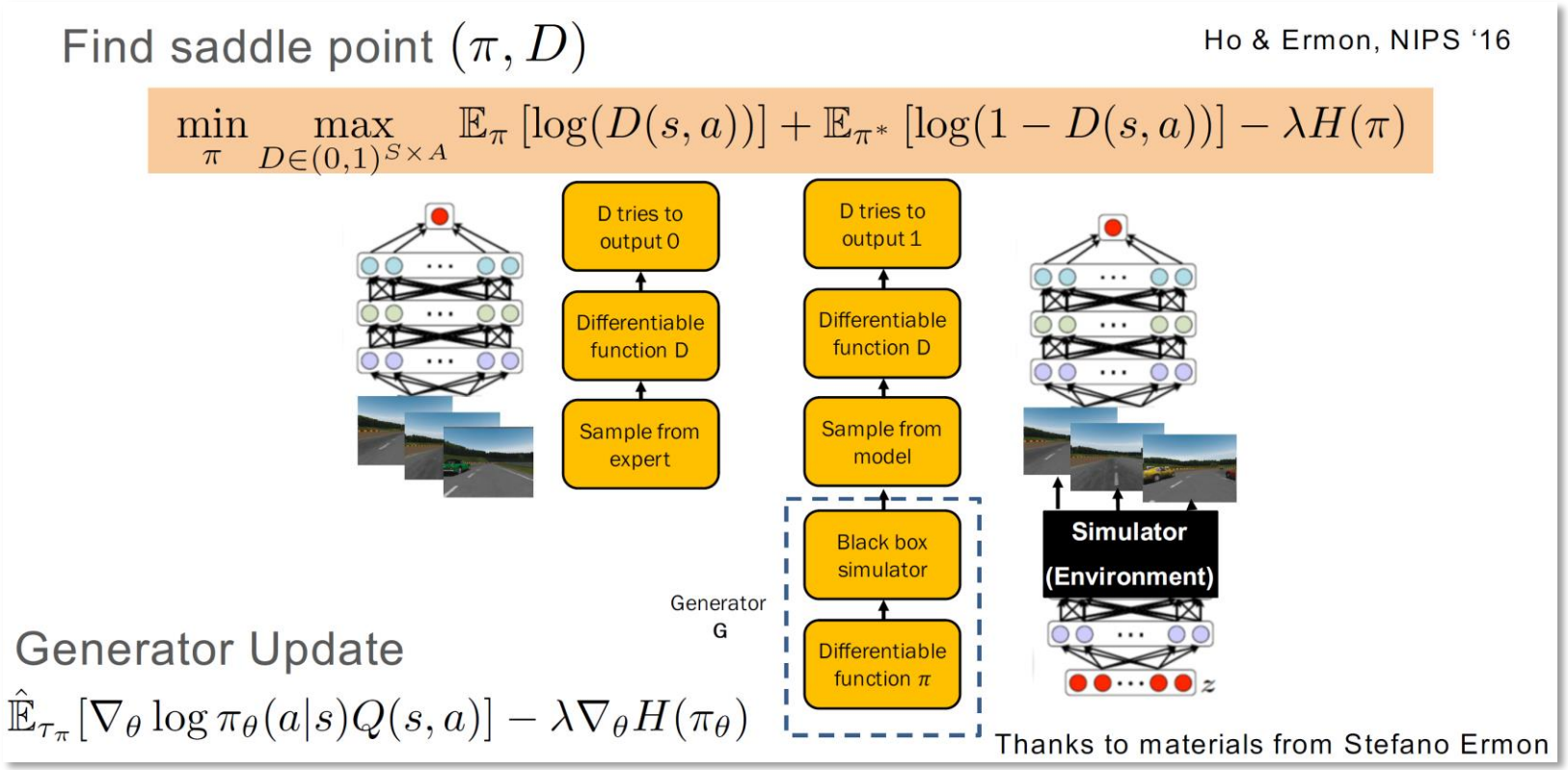
- Challenges in RL with reward engineering → Inverse RL



Imitation Learning, Y. Yue and H. M. Le, ICML 2018

Imitation learning

- Generative Adversarial Imitation Learning



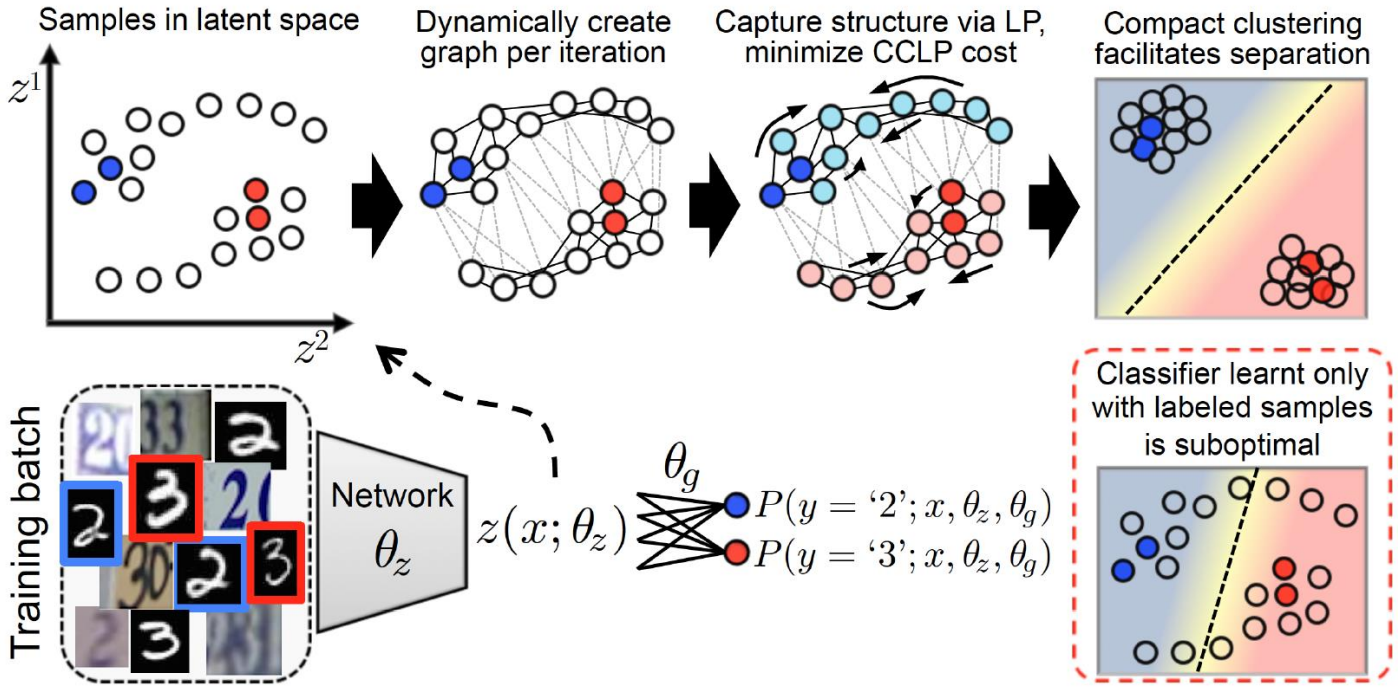
Imitation Learning, Y. Yue and H. M. Le, ICML 2018

Understanding latent spaces of generative models

- Conference
  - Semi-Supervised Learning via Compact Latent Space Clustering
  - Learning Independent Causal Mechanisms
  - Disentangling by Factorising
- Workshop
  - Towards learning with limited labels: Equivariance, Invariance, and Beyond
  - Theoretical Foundations and Applications of Deep Generative Models

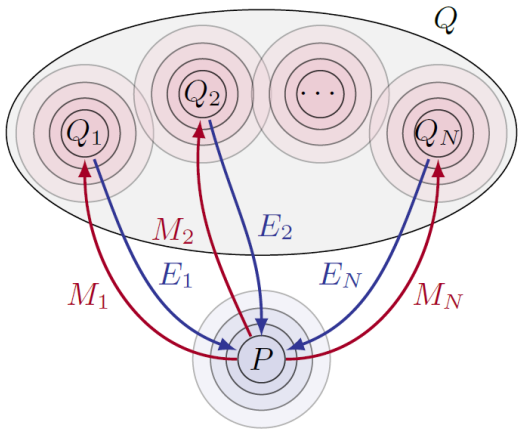
Graph, latent space, and clustering

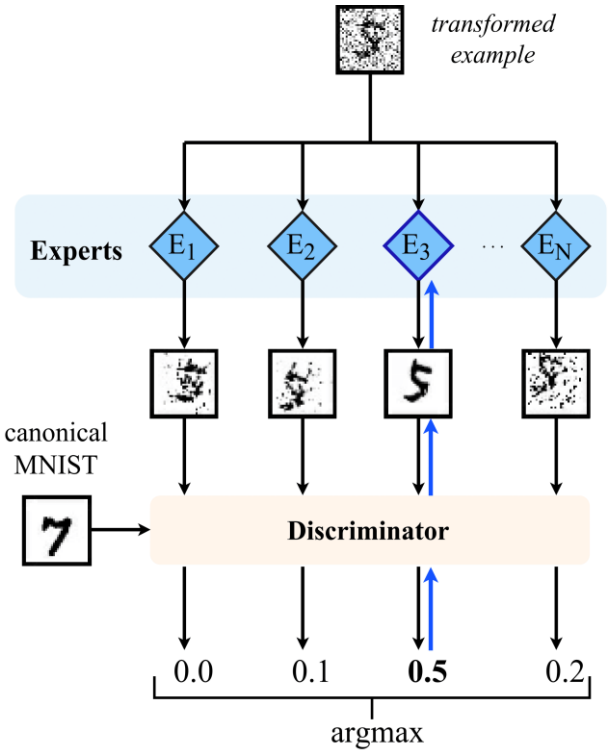
- *Semi-Supervised Learning via Compact Latent Space Clustering*, Konstantinos Kamnitsas et al.
- Method
  1. Dynamically construct a graph in the latent space,
  2. Propagate labels to capture the manifold's structure,
  3. And regularize it to form a single, compact cluster per class.



Disentangling latent space

- *Learning Independent Causal Mechanisms*, Giambattista Parascandolo et al.
  - At training time, an expert transforms a data back to a canonical one.
  - Unsupervised and modular (many experts)
  - A discriminator judges experts
  - Training parallelized across experts





Disentangling latent space

- *Disentangling by Factorising*, Hyunjik Kim and Andriy Mnih.
  - Suggests an objective that directly encourages a factorial latent distribution $q(z) = \Pi_i \, q(z_i)$ and introduce FactorVAE.
  - Comparison to β-VAE and InfoGAN, which are popular generative models for disentangling latent space.
- *Towards disentangling underlying explanatory factors*, Yoshua Bengio
  - *Towards learning with limited labels: Equivariance, Invariance*, and Beyond, ICML 2018 Workshop
  - If the latent space of a generative model is factorized, linear interpolation in the latent space generates good images all along.

Theoretical Foundations and Applications of Deep Generative Models

- *Capturing Dependencies Implicitly,* Yoshua Bengio.
  - Instead of MLE loss in the pixel space, which greatly penalize for not putting a probability distribution on a single data point, we need a loss in the latent space.
- *Interpretable and Semantics-aware Generative Models*, Pushmeet Kohli.
  - Replacing the decoder of an autoencoder with graphics engines for interpretability.
- *Editing is Easier than Generation*, Percy Liang.
  - Learning local transformations represented by matrix Lie group from unlabeled data.
  - Text generation not from scratch but using transformations.
- *Learning hierarchical generative models with structured representations*, Honglak Lee.
  - Hierarchical text-to-image synthesis, text → box → mask → pixel.

Plan

- Highlighted topics
  - Security of ML
  - Fair ML
  - Bayesian Inference
  - Theory of Deep Learning
- Interesting topics
  - Geometry and Deep Learning
  - Replacing Heuristics with Machine Learning
  - Understanding Latent Spaces of Generative models
- **Other topics**

Other topics

- Hierarchical learning
- World model
  - *Building Machines that Learn and Think Like People*, J. Tenenbaum, ICML 2018 Keynote speech.
- AutoML workshop
  - *Automating machine learning*, Zoubin Ghahramani.
    - https://www.automaticstatistician.com

Thank you!

- ML^2, a machine learning research group at KC, is looking for:
  - ML research scientist
  - ML research engineer
  - Contact me by `chan.y.park@kct.co.kr`
-  , a start-up for product personalization, is looking for:
  - Data scientist
  - Data engineer
  - Or the two in one person. ☺
  - Reach me by `chan@morulabs.com`
- https://www.linkedin.com/in/chan-youn-park/
- Any question or collaboration is always welcome!