Deep Learning Limitations and New Frontiers

Ava Soleimany

MIT 6.S191

January 26, 2022

T-shirts! Today!

MIT 6.S191

Lecture Schedule

Intro to Deep Learning
Lecture 1
Jan. 24, 2022
[Slides] [Video] *coming soon!*

Deep Sequence Modeling
Lecture 2
Jan. 24, 2022
[Slides] [Video] *coming soon!*

Intro to TensorFlow; Music Generation
Software Lab 1
[Code] *coming soon!*

Deep Computer Vision
Lecture 3
Jan. 25, 2022
[Slides] [Video] *coming soon!*

Deep Generative Modeling
Lecture 4
Jan. 25, 2022
[Slides] [Video] *coming soon!*

De-biasing Facial Recognition Systems
Software Lab 2
[Paper] [Code] *coming soon!*

Deep Reinforcement Learning
Lecture 5
Jan. 26, 2022
[Slides] [Video] *coming soon!*

Limitations and New Frontiers
Lecture 6
Jan. 26, 2022
[Slides] [Video] *coming soon!*

Learning End-to-End Self-Driving Control
Software Lab 3
[Code] *coming soon!*

Autonomous Driving with LiDAR
Lecture 7
Jan. 27, 2022
[Info] [Slides] [Video] *coming soon!*

Uncertainty in Deep Learning
Lecture 8
Jan. 27, 2022
[Info] [Slides] [Video] *coming soon!*

Final Project
Work on final projects

Guest Lecture
Lecture 9
Jan. 28, 2022
[Info] [Slides] [Video] *coming soon!*

Speech Recognition
Lecture 10
Jan. 28, 2022
[Info] [Slides] [Video] *coming soon!*

Project Competition
Project pitches and final awards!

- Lab submission: 1/27/22
- Paper review: 1/28/22
- Final projects: 1/28/22

Labs and Prizes

Lab 1: Music Generation

Beats Headphones

Lab 2: Computer Vision

24" HD Display Monitor

Lab 3: Reinforcement Learning

Quadcopter Drone

+ Deploy your model on a real self-driving car

Lab submission: 1/27/22

Final Class Project

Option 1: Proposal Presentation

- At least 1 registered student to be prize eligible
- Present a novel deep learning research idea or application
- 3 minutes (strict)
- Presentations on **Friday, Jan 28**
- Submit groups by **Wed 1/26 11:59pm ET** to be eligible
- Submit slides by **Thu 1/27 11:59pm ET** to be eligible
- Instructions: bit.ly/3qOQEuG

- Judged by a panel of judges
- Top winners are awarded:

NVIDIA 3080 GPU

4x Smartwatches

3x Display Monitors

MIT 6.S191

Final Class Project

Option 1: Proposal Presentation

- At least 1 registered student to be prize eligible
- Present a novel deep learning research idea or application
- 3 minutes (strict)
- Presentations on Friday, Jan 29
- Submit groups by Wednesday 11:59pm ET to be eligible
- Submit slide by Thursday 11:59pm ET to be eligible
- Instructions:

Option 2: Write a 1-page review of a deep learning paper

- Grade is based on clarity of writing and technical communication of main ideas
- Due Fri Jan 28 3:59pm ET

MIT 6.S191

Up Next: Guest Lectures


Omer Keilaf
Amir Day
Innoviz

Jasper Snoek
Google

Anima Anandkumar
NVIDIA
Caltech


Miguel Jette
Jenny Drexler
RevAI

So far in 6.S191…

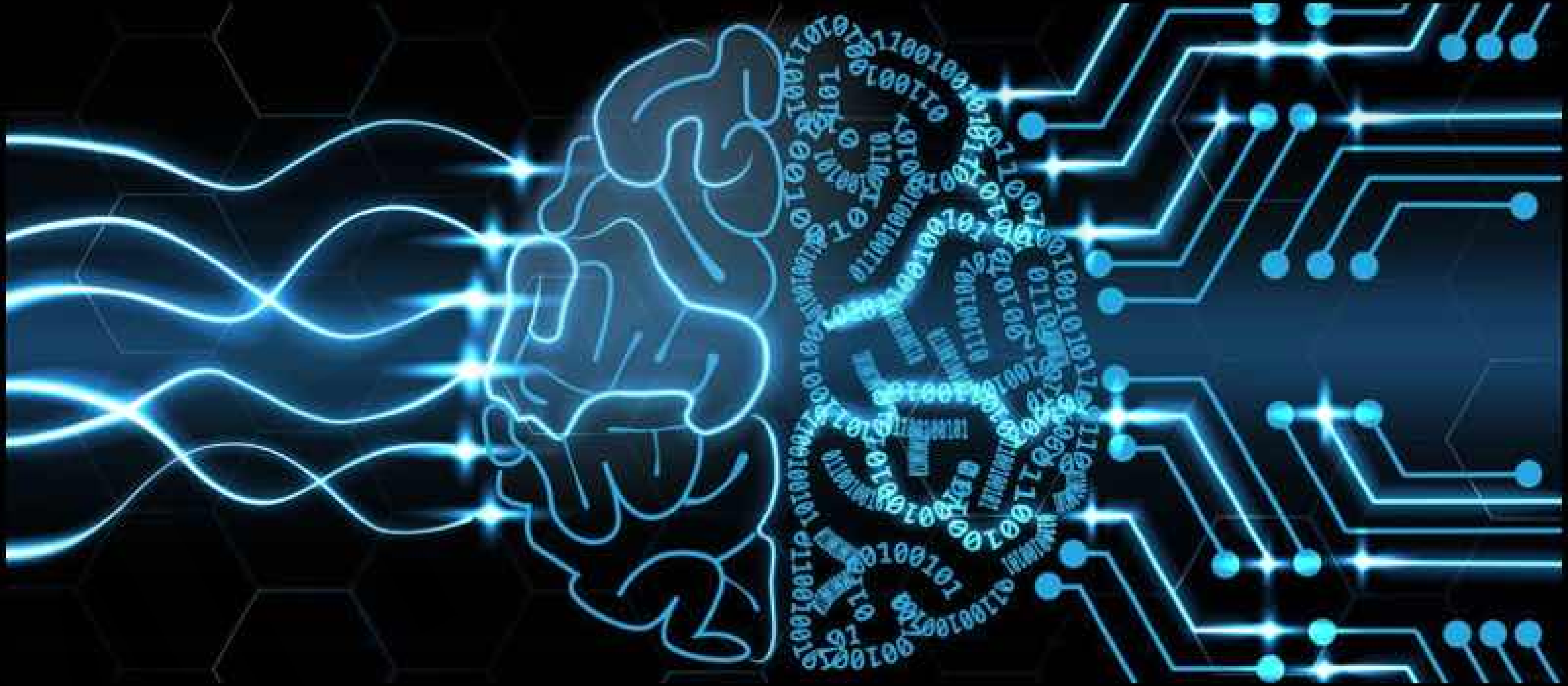
MIT 6.S191

The Rise of Deep Learning

So far in 6.S191…

Data
- Signals
- Images
- Sensors

…



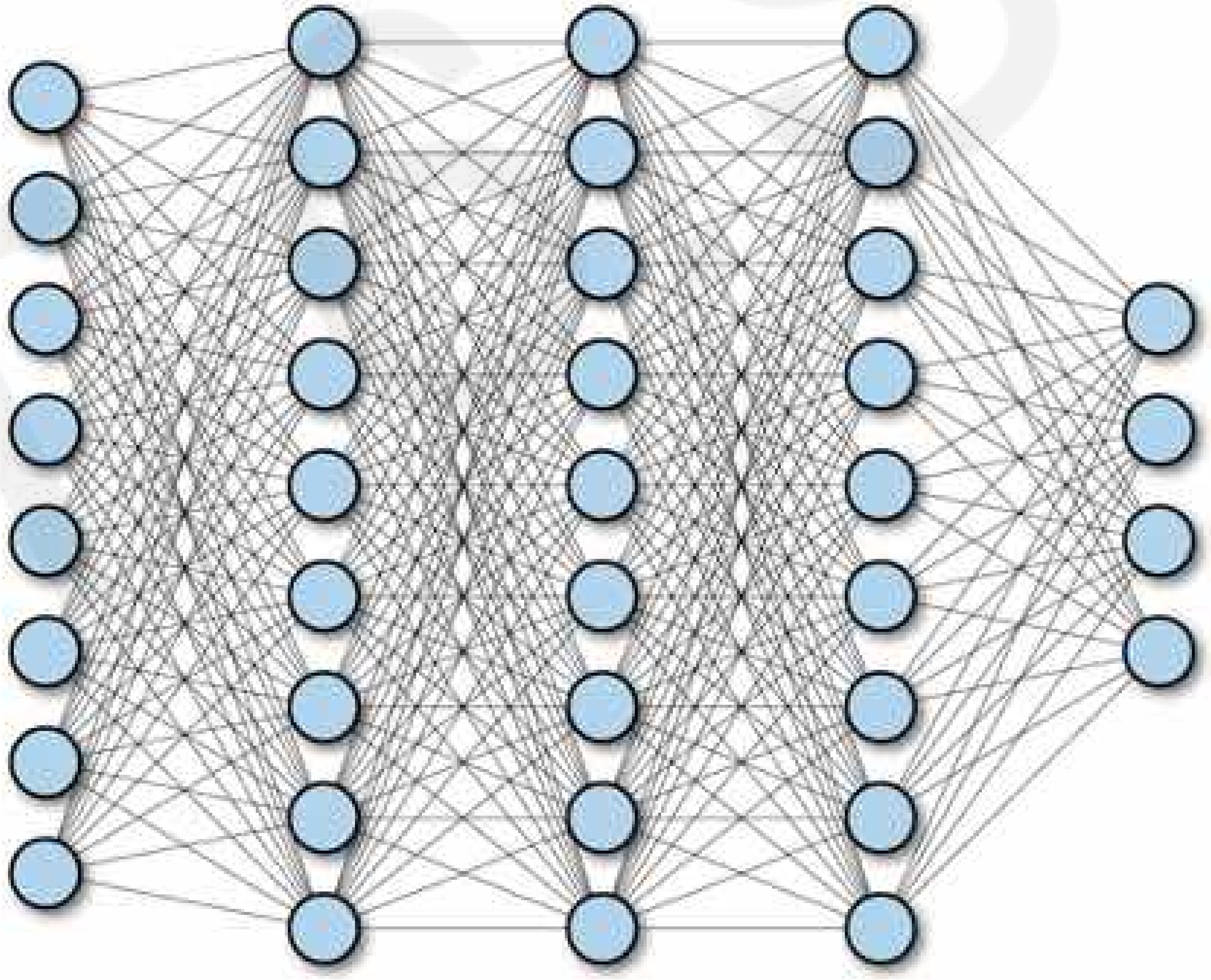
Decision
- Prediction
- Detection
- Action

…

Power of Neural Nets

Universal Approximation Theorem

A feedforward network with a single layer is sufficient to approximate, to an arbitrary precision, any continuous function.



MIT S191

Power of Neural Nets

Universal Approximation Theorem

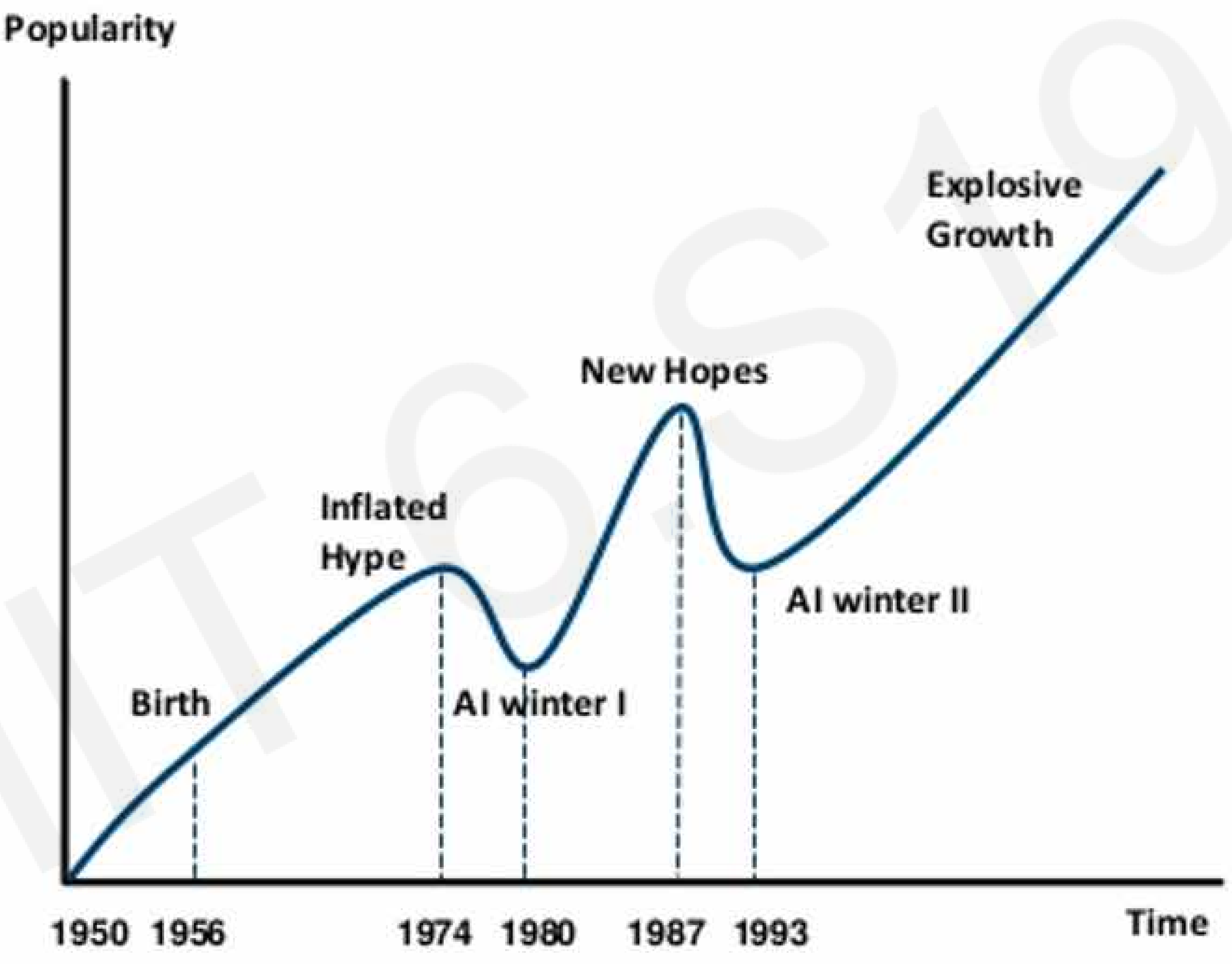
A feedforward network with a single layer is sufficient to approximate, to an arbitrary precision, any continuous function.

Caveats:

- *The number of hidden units may be infeasibly large*
- *The resulting model may not generalize*

MIT 6.S191

Artificial Intelligence "Hype": Historical Perspective



MIT 6.S191

Limitations

MIT 6.S191

Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

dog

banana

dog

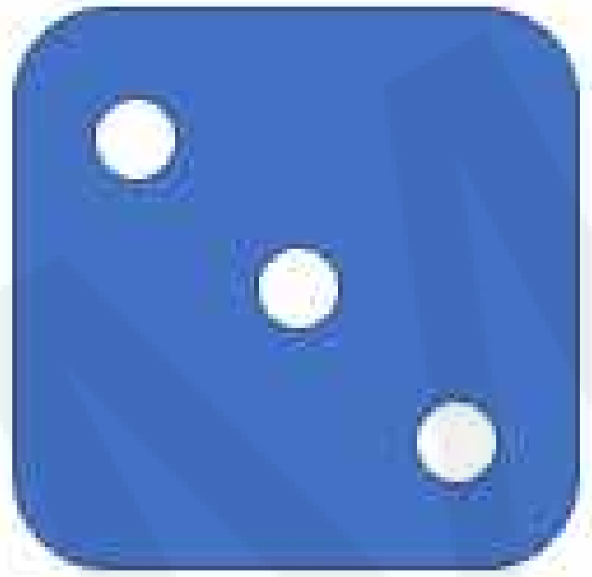
tree

MIT 6.S191

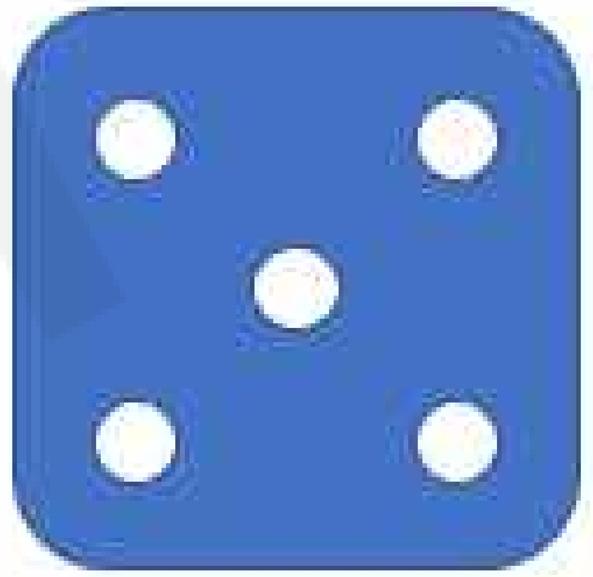
Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

dog

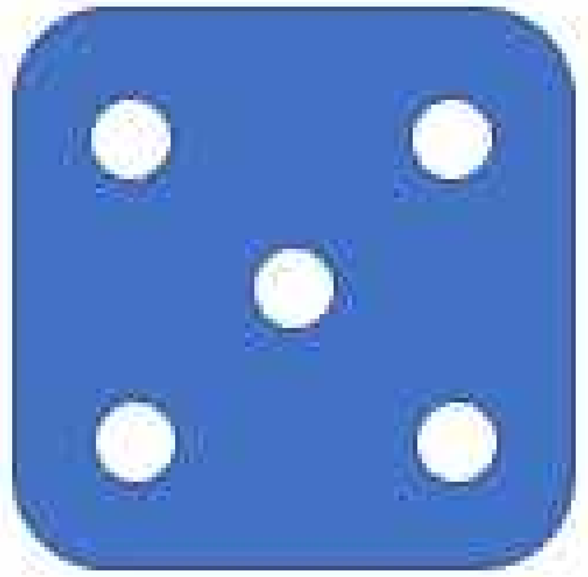


banana



dog

tree



Rethinking Generalization

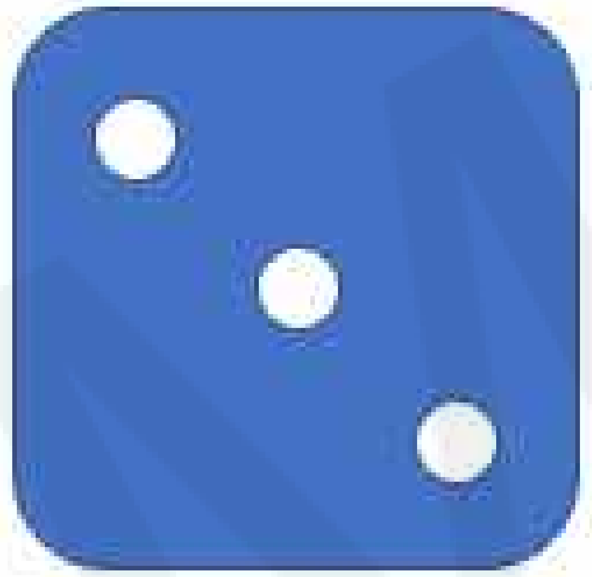
"Understanding Deep Neural Networks Requires Rethinking Generalization"

dog

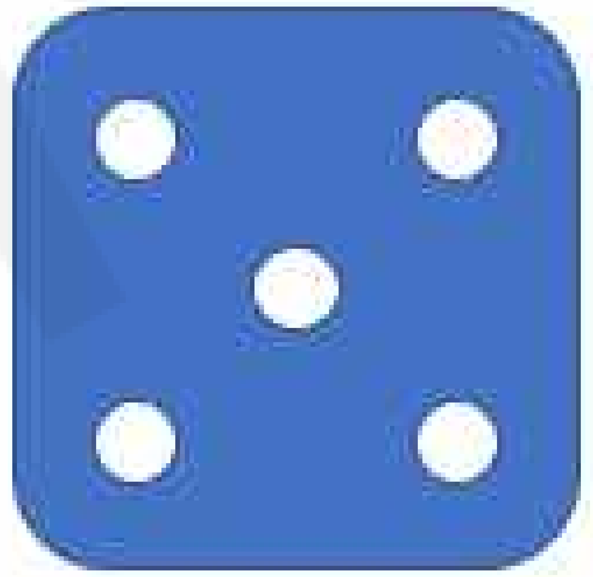
banana

dog

tree

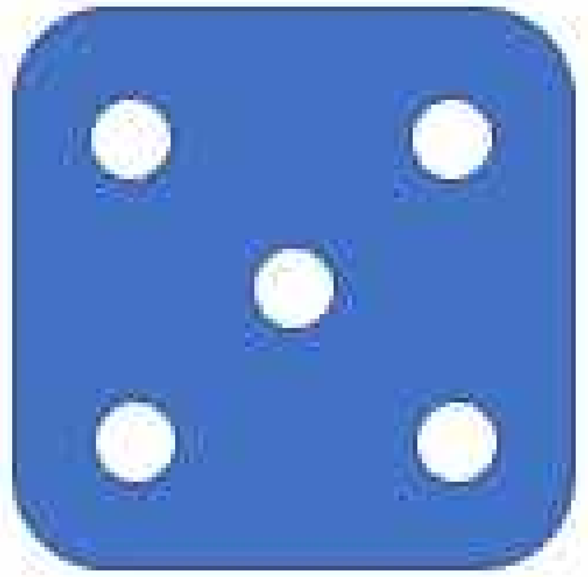


banana



dog

tree



dog

MIT 6.S191

Rethinking Generalization

"Understanding Deep Neural Networks Requires Rethinking Generalization"

~~dog~~

banana

~~banana~~

dog

~~dog~~

tree

~~tree~~

dog

MIT 6.S191

Capacity of Deep Neural Networks

MIT 6.S191

Capacity of Deep Neural Networks

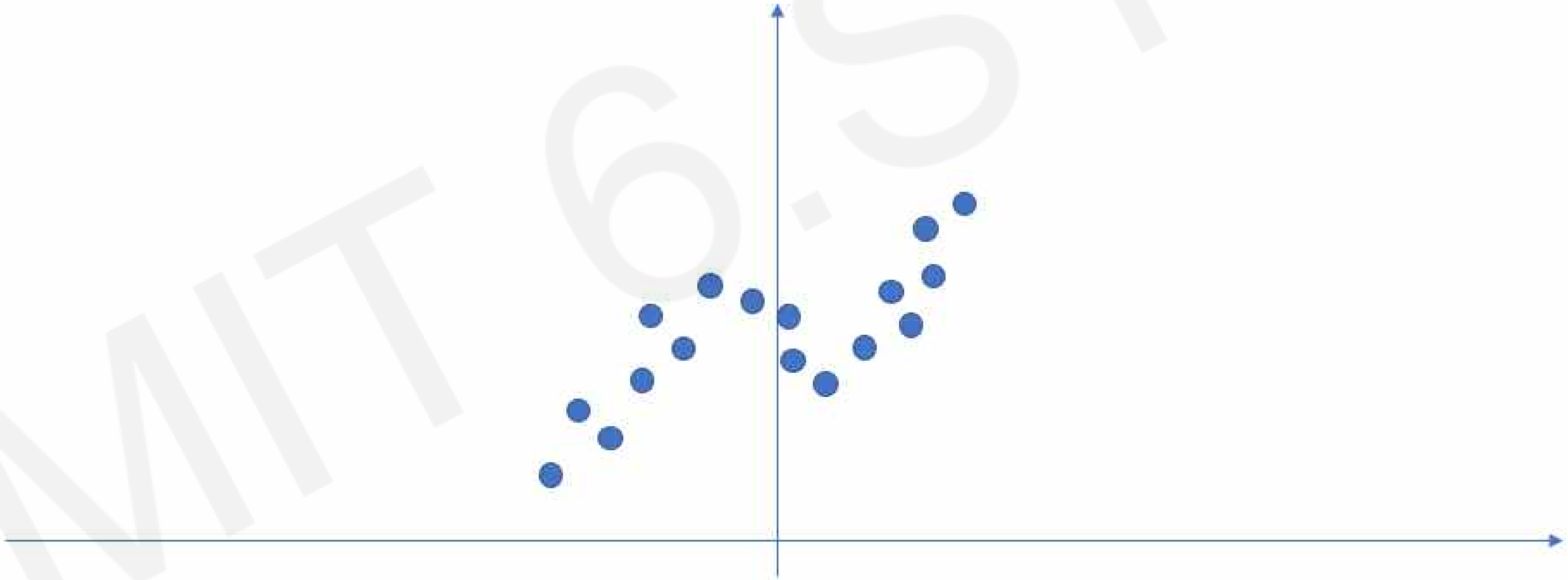
MIT 6.S191

Capacity of Deep Neural Networks

MIT 6.S191

Neural Networks as Function Approximators

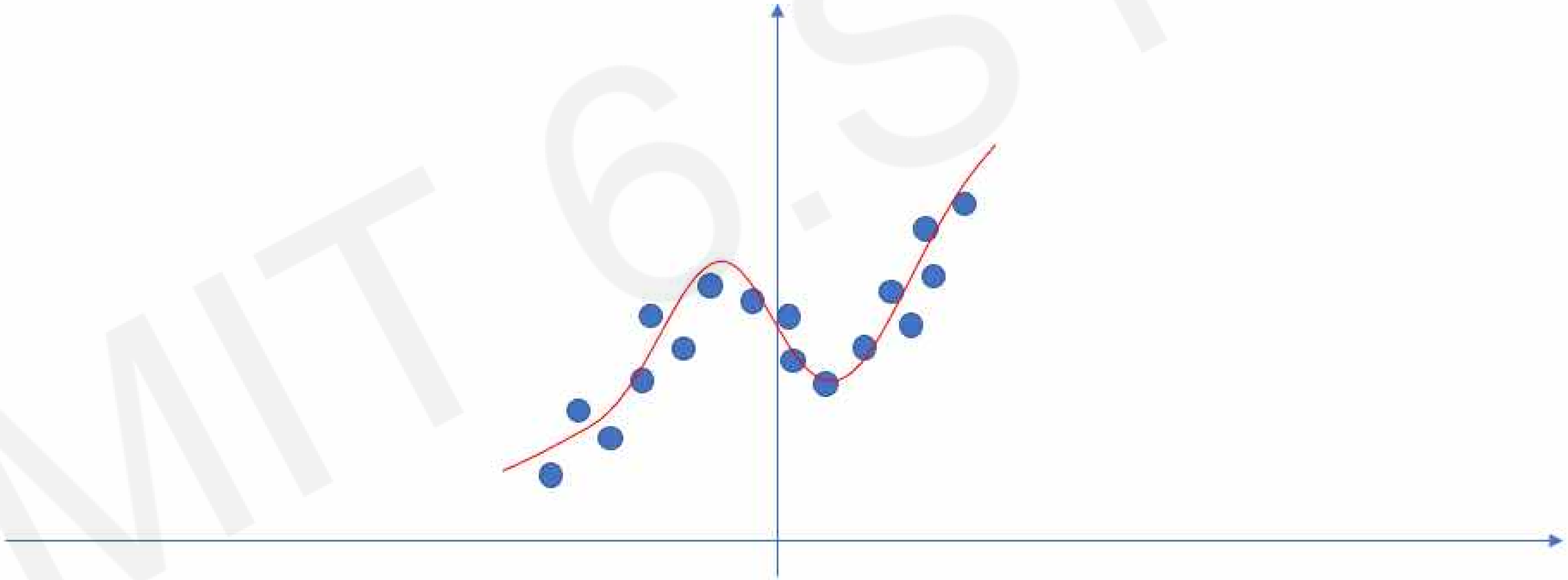
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

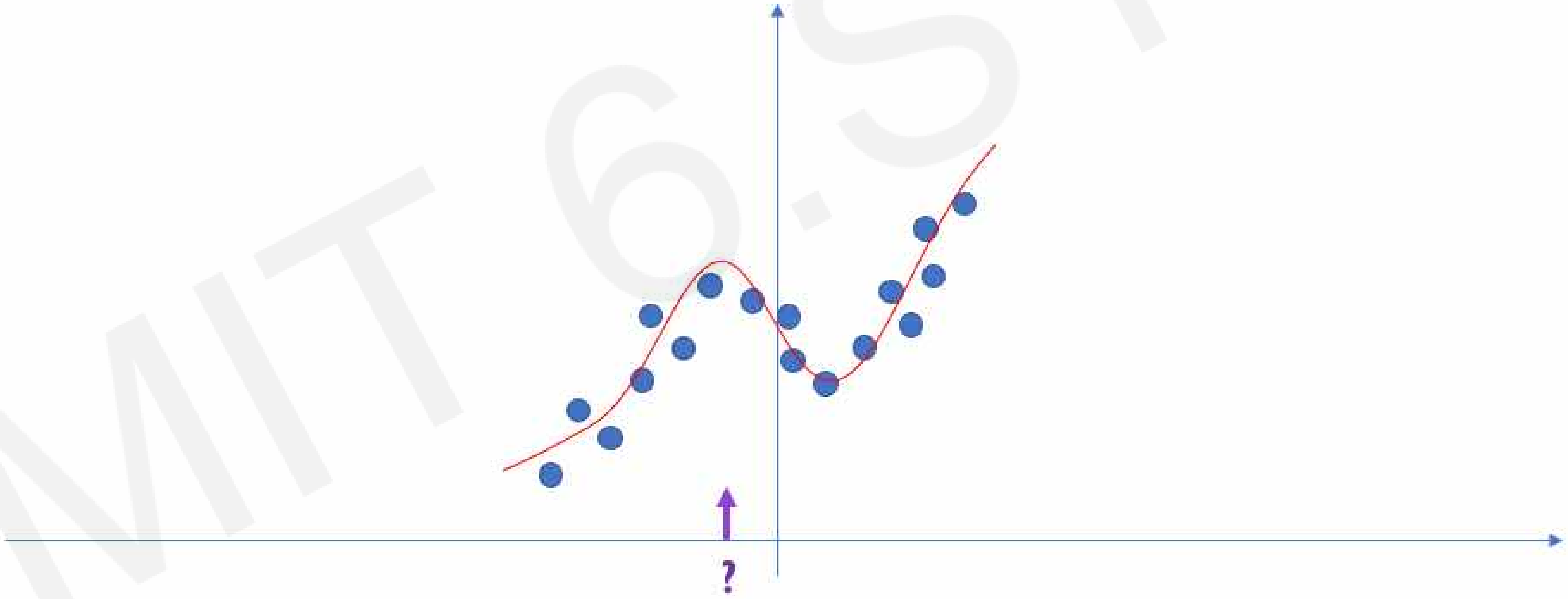
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

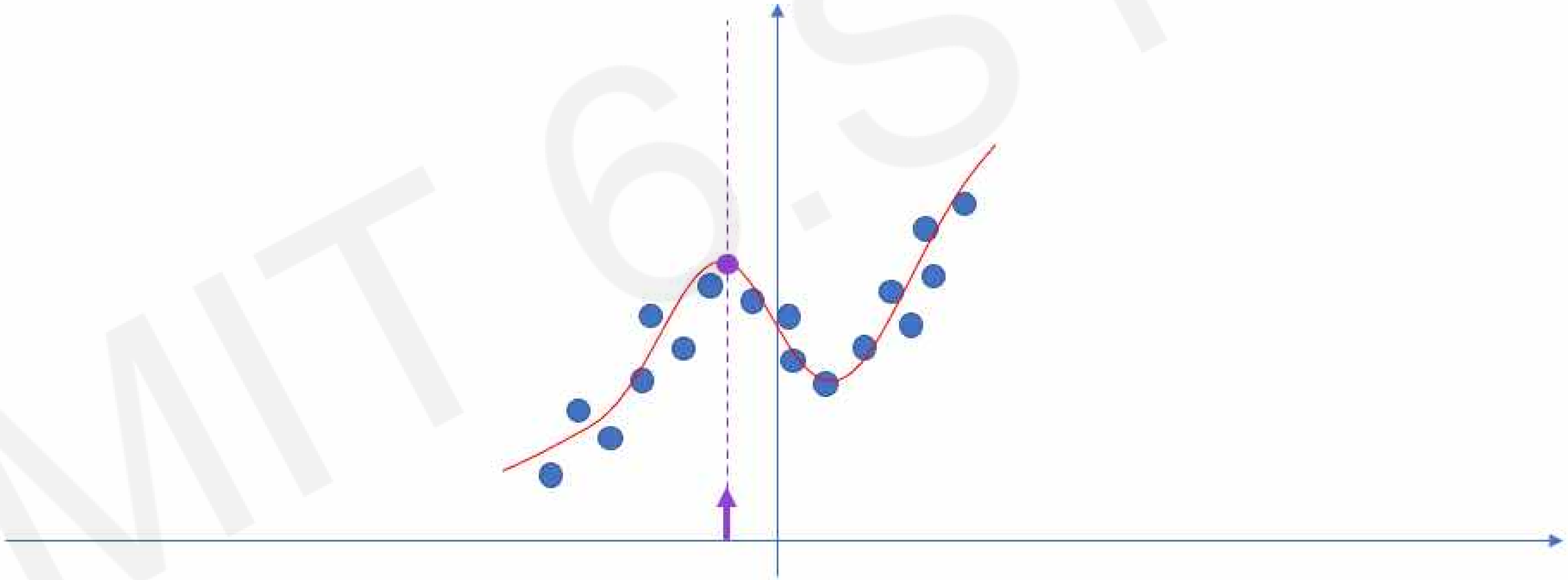
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

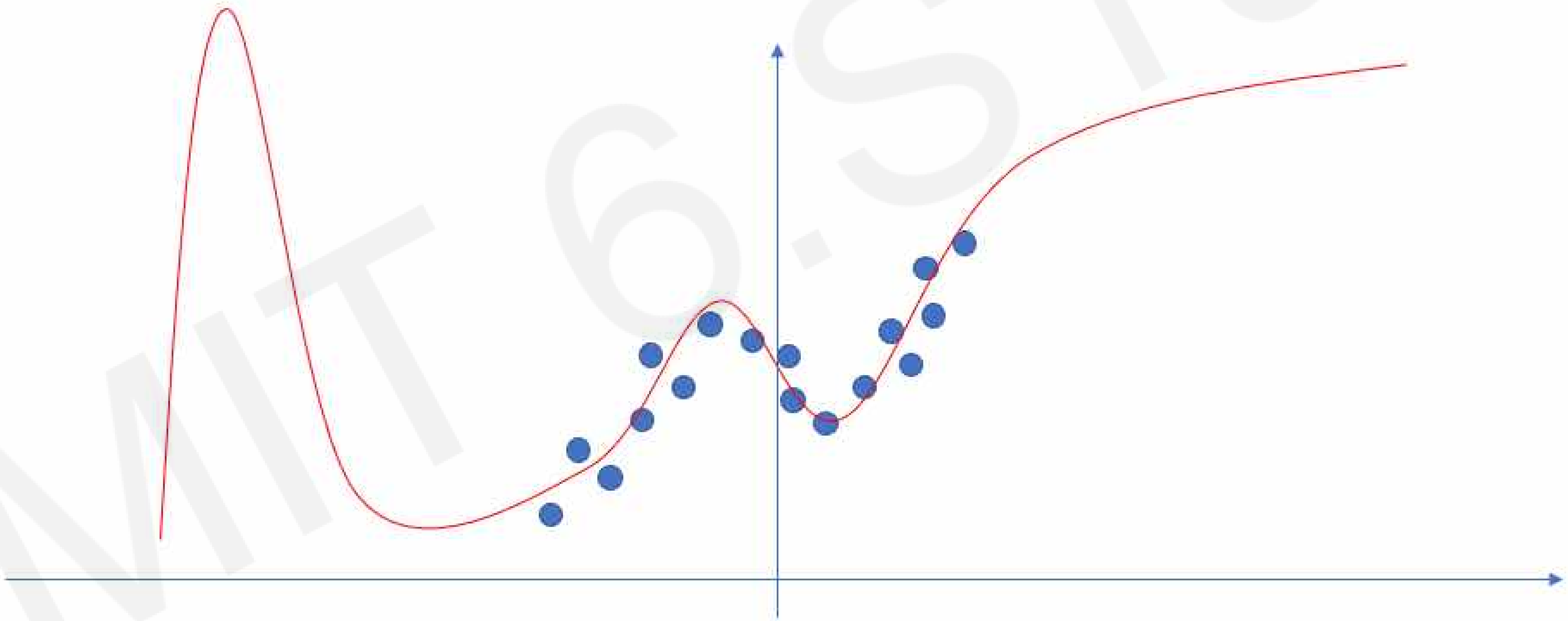
Neural networks are **excellent** function approximators



MIT 6.S191

Neural Networks as Function Approximators

Neural networks are **excellent** function approximators

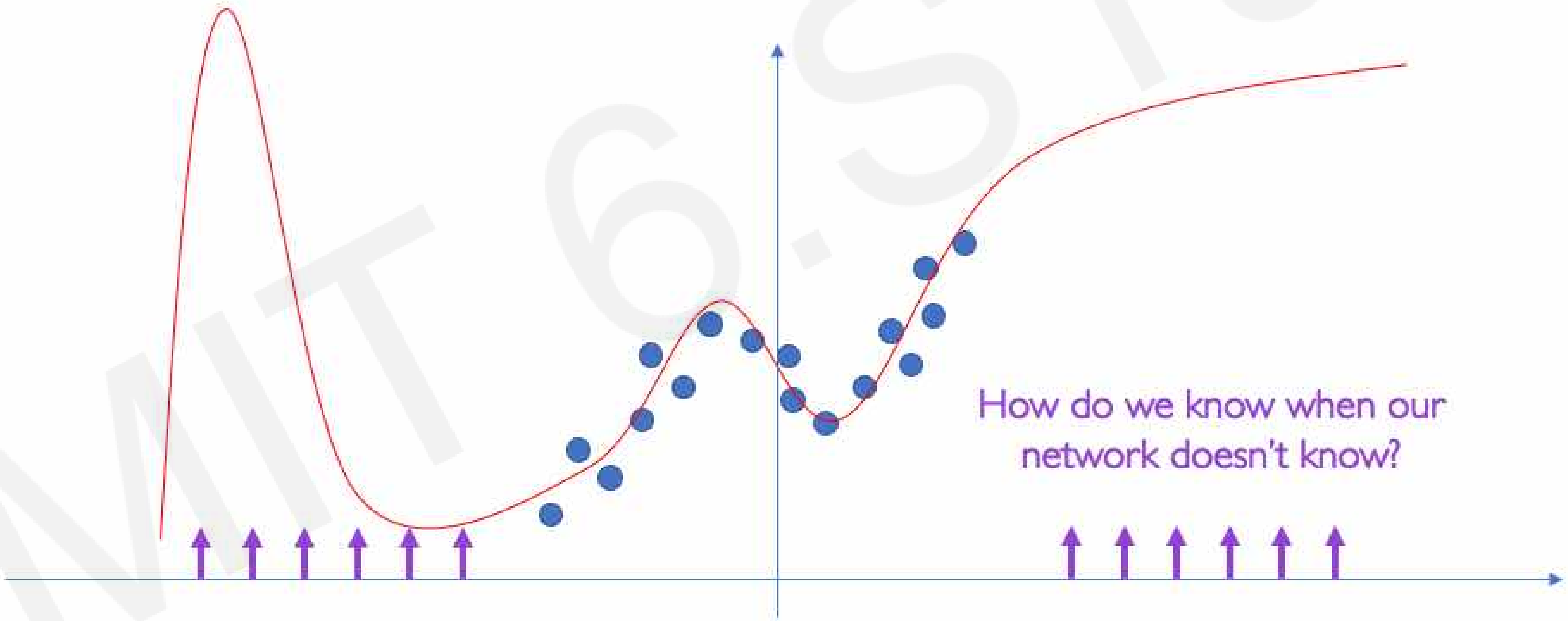


MIT 6.S191

Neural Networks as Function Approximators

Neural networks are **excellent** function approximators

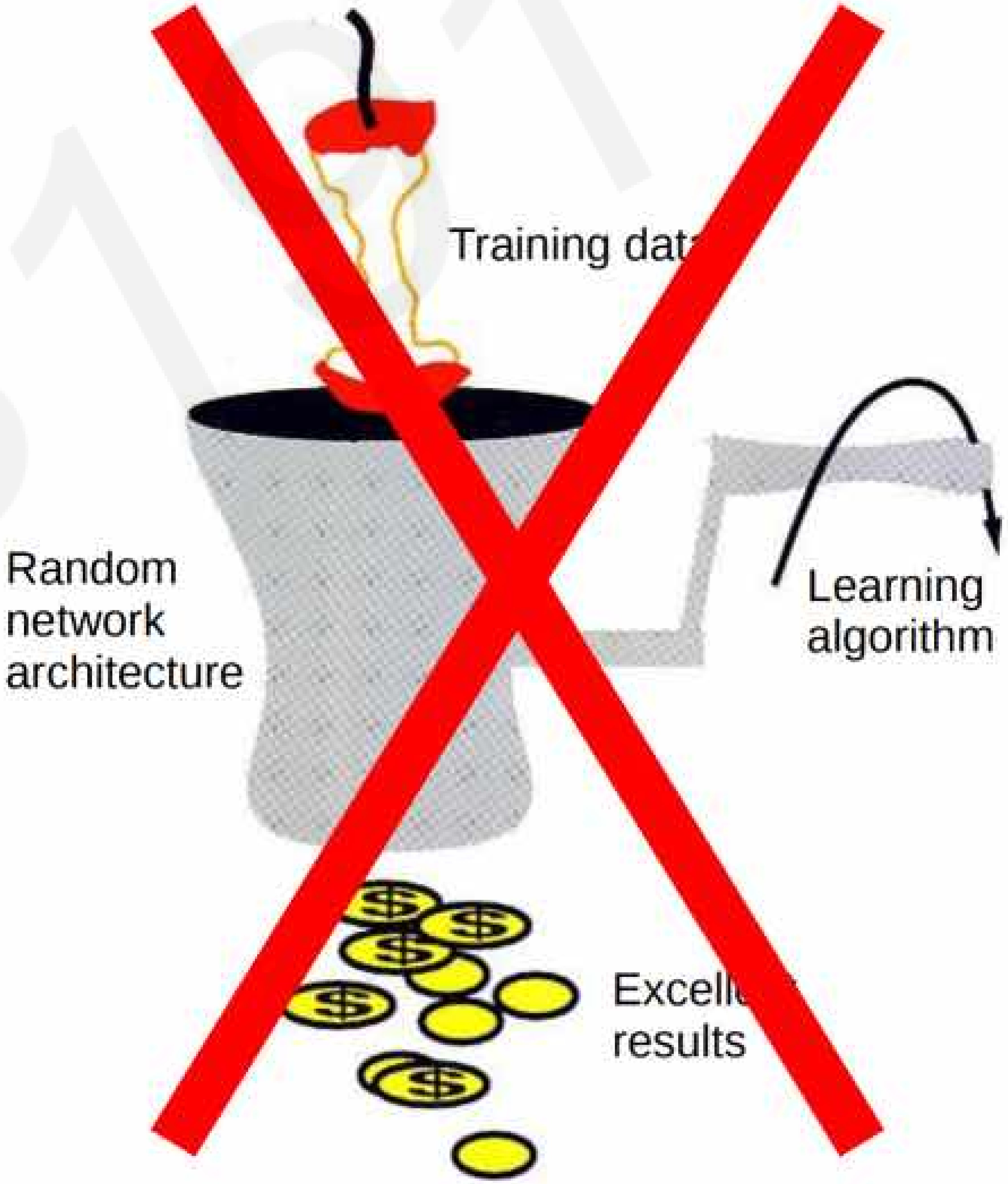
…when they have training data



MIT 6.S191

Deep Learning = Alchemy?

NO!



Neural Network Failure Modes, Part 1

Train network to
colorize BW images.

Why could this be the case?

What Happens During Training…

MIT 6.S191

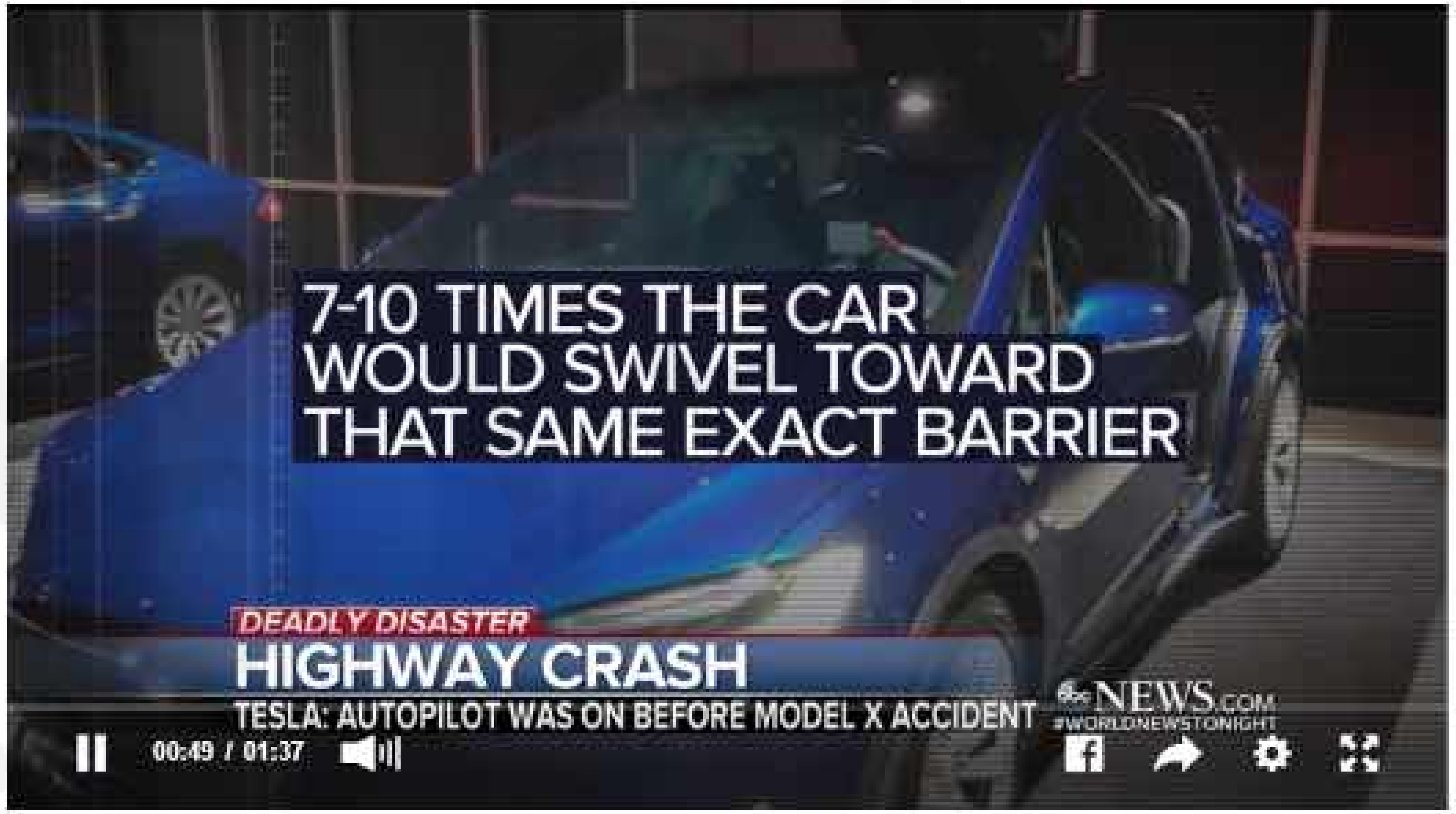
Neural Network Failure Modes, Part II

Tesla car was on autopilot prior to fatal crash in California, company says

The crash near Mountain View, California, last week killed the driver.

By Mark Osborne
March 31, 2018, 1:57 AM • 5 min read

abc NEWS

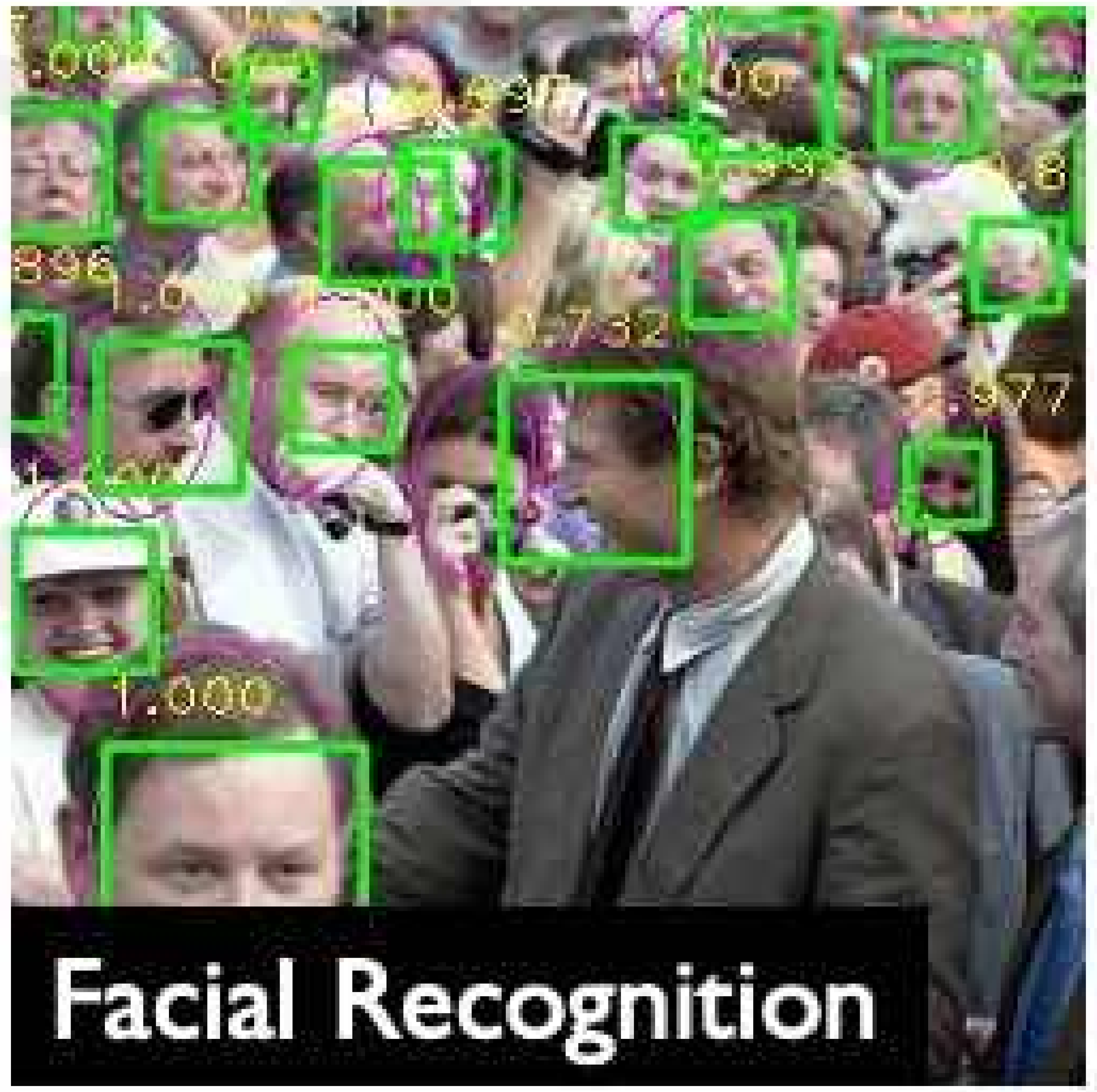


Uncertainty in Deep Learning

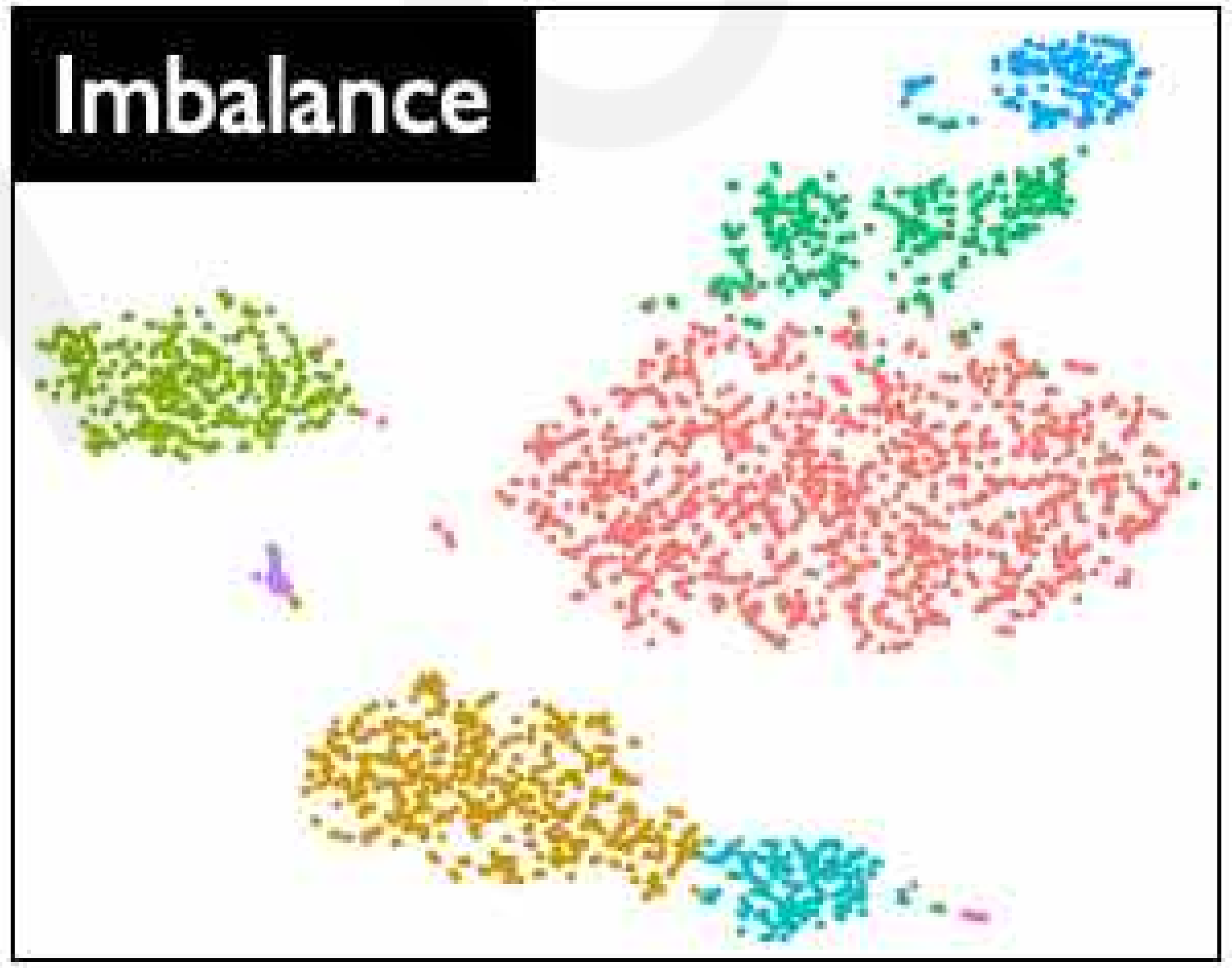
Safety-critical applications


Autonomous Vehicles


Medicine


Facial Recognition

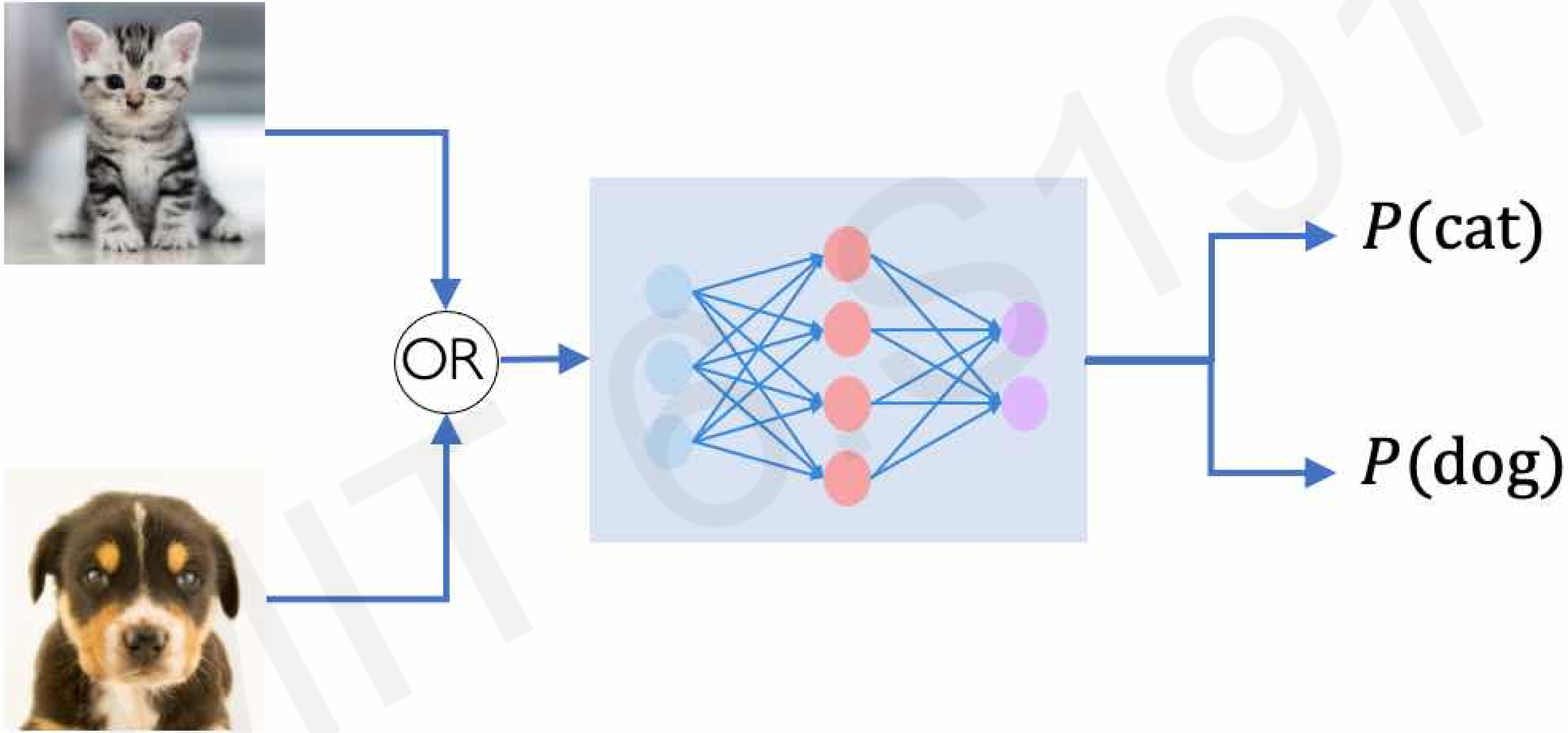
Sparse and/or noisy datasets


Imbalance


Data Noise

6.S191 Guest Lecture

What uncertainties do we need?

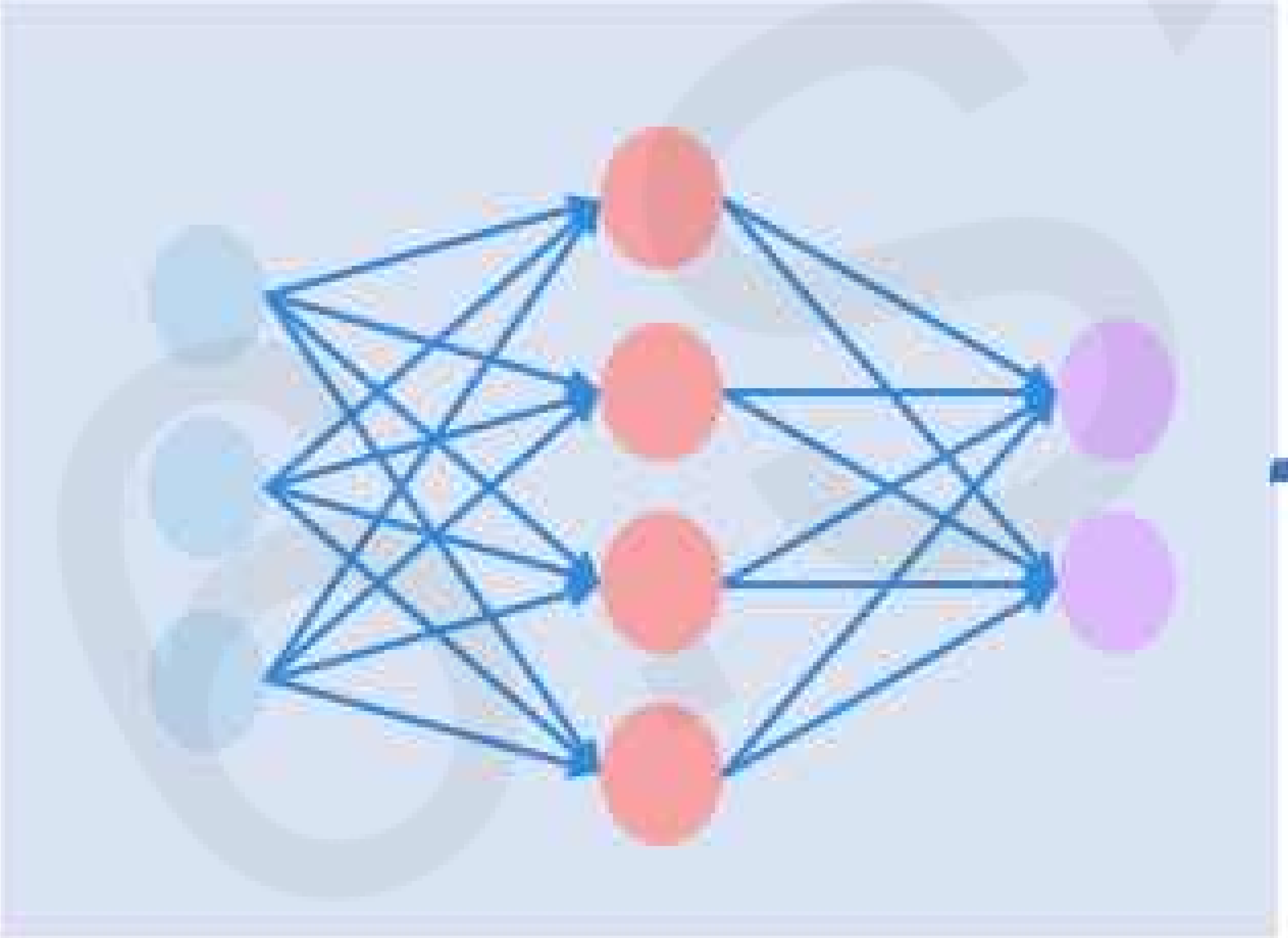


6.S191 Lecture

What uncertainties do we need?

We need uncertainty metrics to assess the noise inherent to the data.

aleatoric uncertainty



$P(\text{cat}) = 0.5$

$P(\text{dog}) = 0.5$

Remember: $P(\text{cat}) + P(\text{dog}) = 1$

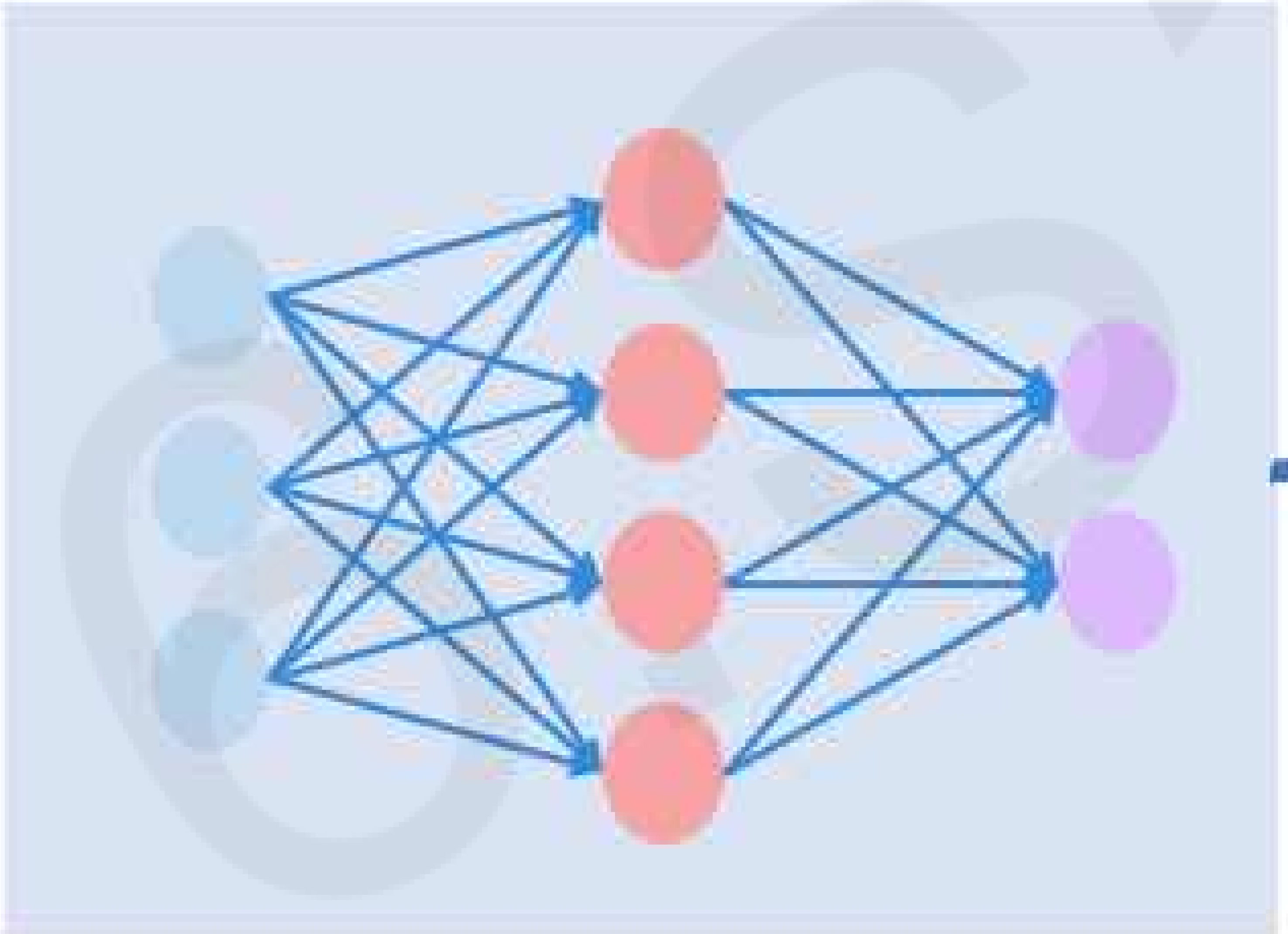
6.S191 Lecture

MIT 6.S191

What uncertainties do we need?

We need uncertainty metrics to assess the network's confidence in its predictions.

epistemic uncertainty



$P(\text{cat}) = 0.2$

$P(\text{dog}) = 0.8$

Remember: $P(\text{cat}) + P(\text{dog}) = 1$

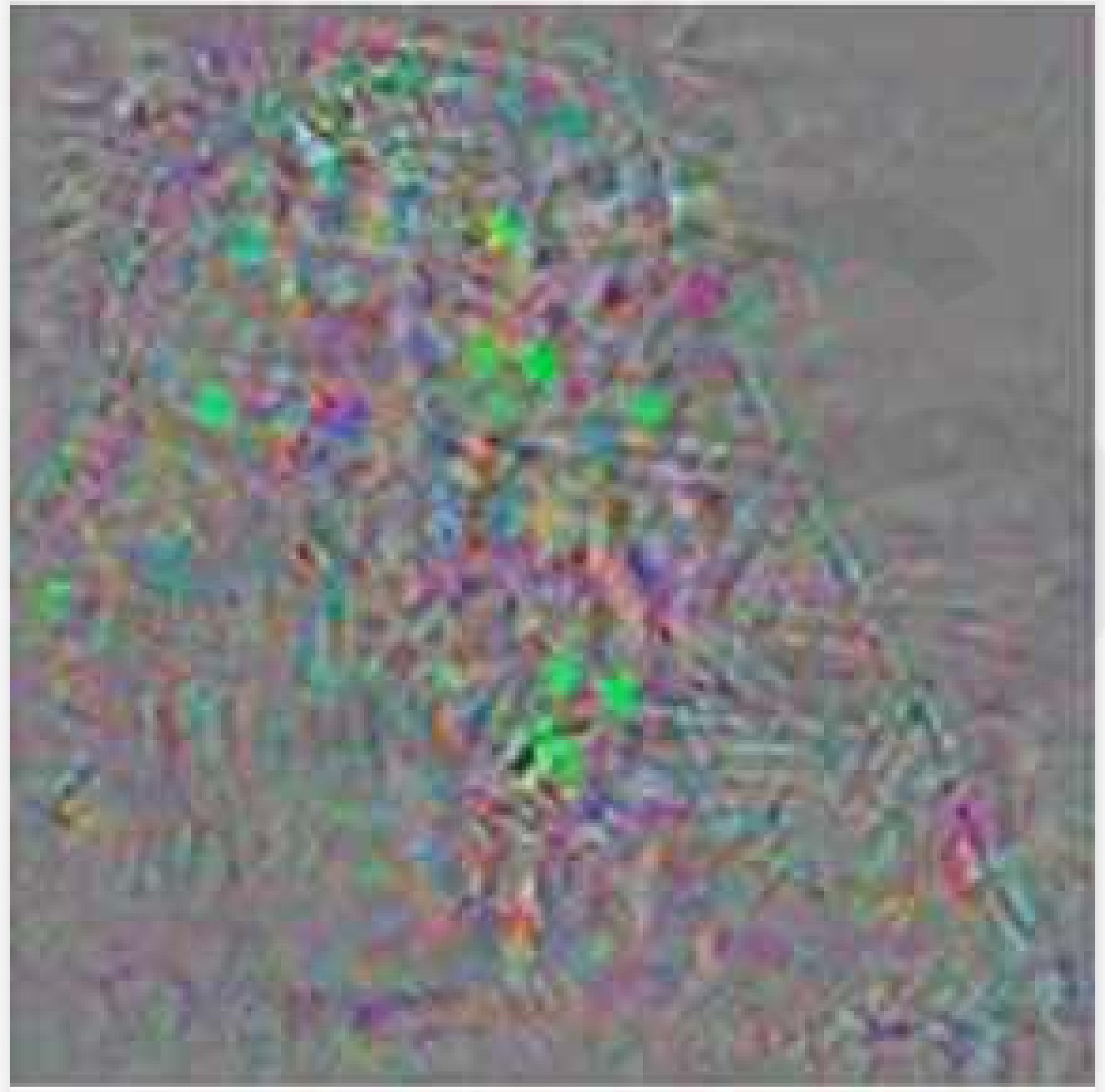
6.S191 Lecture

MIT 6.S191

Neural Network Failure Modes, Part III

Original image

Temple (97%)



Perturbations

Adversarial example

Ostrich (98%)

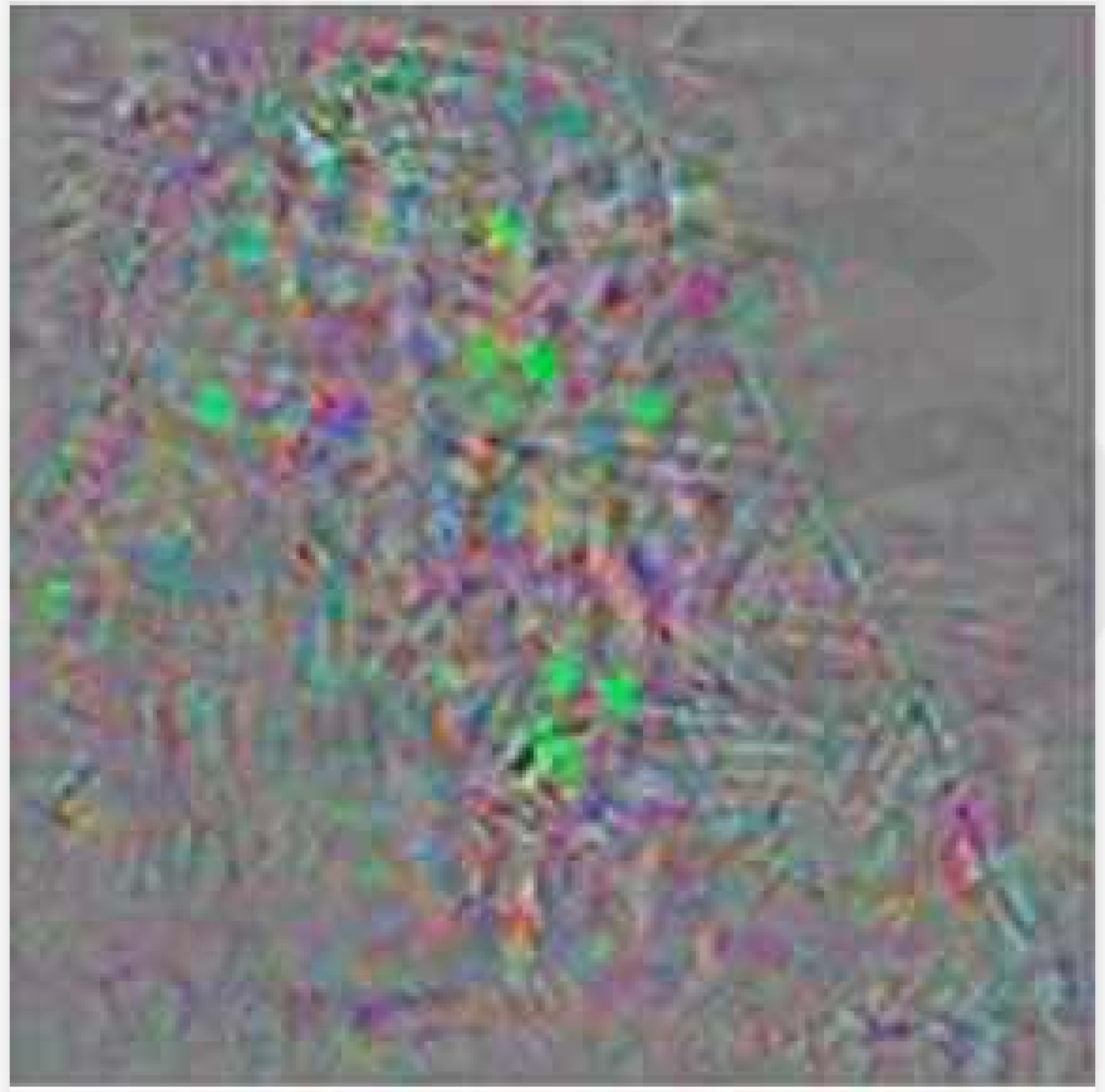
MIT 6.S191

Despois "Adversarial examples" 2017. 1/26/22

Adversarial Attacks on Neural Networks

Original image

Temple (97%)



Perturbations

Adversarial example

Ostrich (98%)

MIT 6.S191

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$W \leftarrow W - \eta \frac{\partial J(W, x, y)}{\partial W}$$

"How does a small change in weights decrease our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$\boxed{W} \leftarrow \boxed{W} - \eta \frac{\partial J(\boxed{W}, x, y)}{\partial \boxed{W}}$$

"How does a small change in weights decrease our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Remember:

We train our networks with gradient descent

$$\boxed{W} \leftarrow \boxed{W} - \eta \frac{\partial J(\boxed{W}, \boxed{x, y})}{\partial \boxed{W}}$$

Fix your image $\boldsymbol{x}$, and true label $\boldsymbol{y}$

"How does a small change in weights decrease our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

"How does a small change in the input increase our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

"How does a small change in the input increase our loss"

MIT 6.S191

Adversarial Attacks on Neural Networks

Adversarial Image:

Modify image to increase error

$$x \leftarrow x + \eta \frac{\partial J(W, x, y)}{\partial x}$$

Fix your weights θ, and true label y

"How does a small change in the input increase our loss"

MIT 6.S191

Synthesizing Robust Adversarial Examples

Algorithmic Bias

Overcoming Racial Bias In AI Systems And Startlingly Even In AI Self-Driving Cars

Racial bias in a medical algorithm favors white patients over sicker black patients

AI expert calls for end to UK use of 'racially biased' algorithms

AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators

Gender bias in AI: building fairer algorithms

Bias in AI: A problem recognized but still unresolved

Amazon, Apple, Google, IBM, and Microsoft worse at transcribing black people's voices than white people's with AI voice recognition, study finds

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn't found one.

The Week in Tech: Algorithmic Bias Is Bad. Uncovering It Is Good.

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

Artificial Intelligence has a gender bias problem – just ask Siri

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

6.S191 Lab

Neural Network Limitations…

- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Difficult to **encode structure** and prior knowledge during learning
- **Finicky to optimize**: non-convex, choice of architecture, learning parameters
- Often require **expert knowledge** to design, fine tune architectures

MIT 6.S191

Neural Network Limitations...

- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Difficult to **encode structure** and prior knowledge during learning
- **Finicky to optimize**: non-convex, choice of architecture, learning parameters
- Often require **expert knowledge** to design, fine tune architectures

6.S191 Lab + Lecture

MIT 6.S191

Neural Network Limitations…

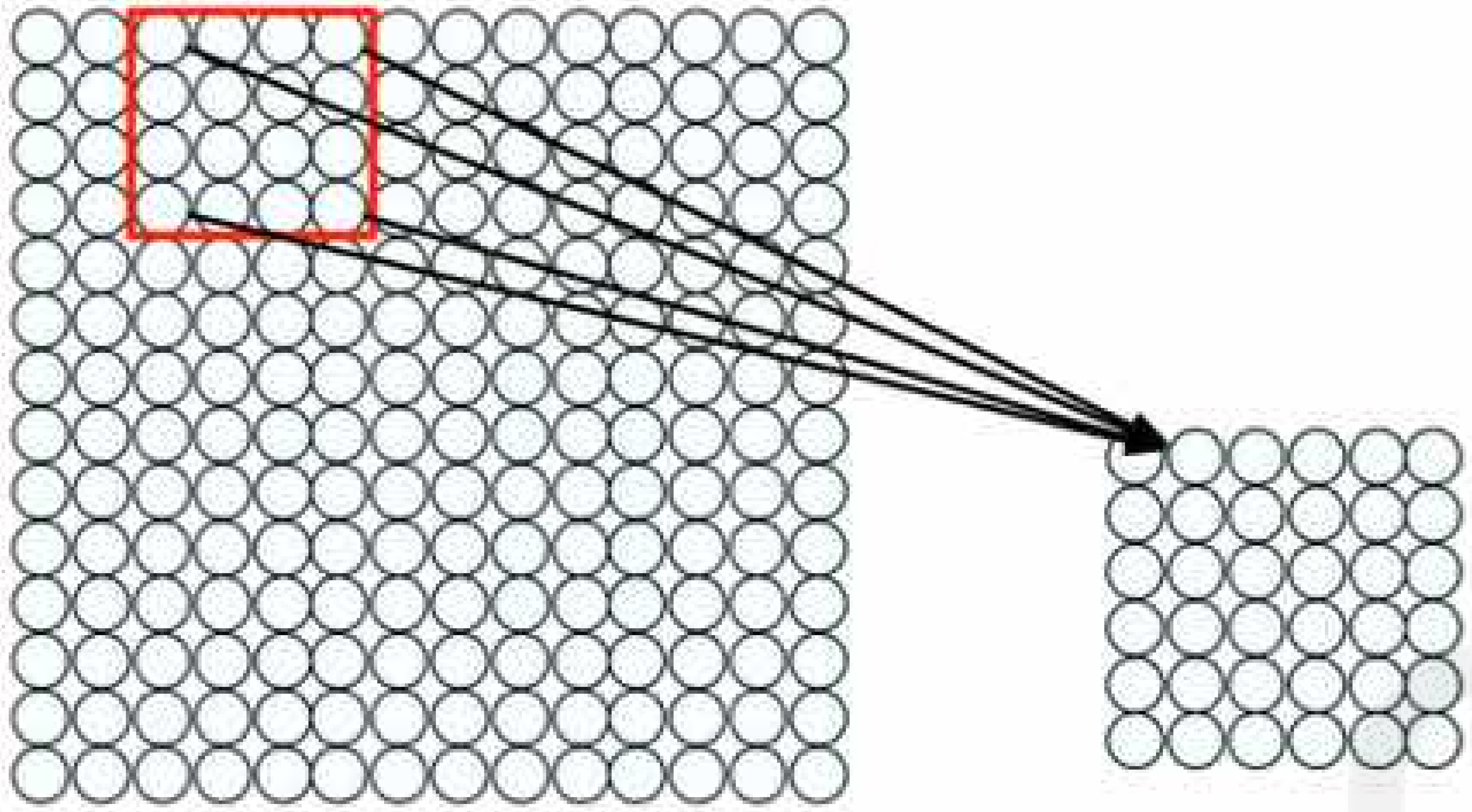
- Very **data hungry** (eg. often millions of examples)
- **Computationally intensive** to train and deploy (tractably requires GPUs)
- Easily fooled by **adversarial examples**
- Can be subject to **algorithmic bias**
- Poor at **representing uncertainty** (how do you know what the model knows?)
- Uninterpretable **black boxes**, difficult to trust
- Difficult to **encode structure** and prior knowledge during learning
- **Finicky to optimize**: non-convex, choice of architecture, learning parameters
- Often require **expert knowledge** to design, fine tune architectures

MIT 6.S191

New Frontiers 1:
Encoding Structure into Deep Learning

MIT 6.S191

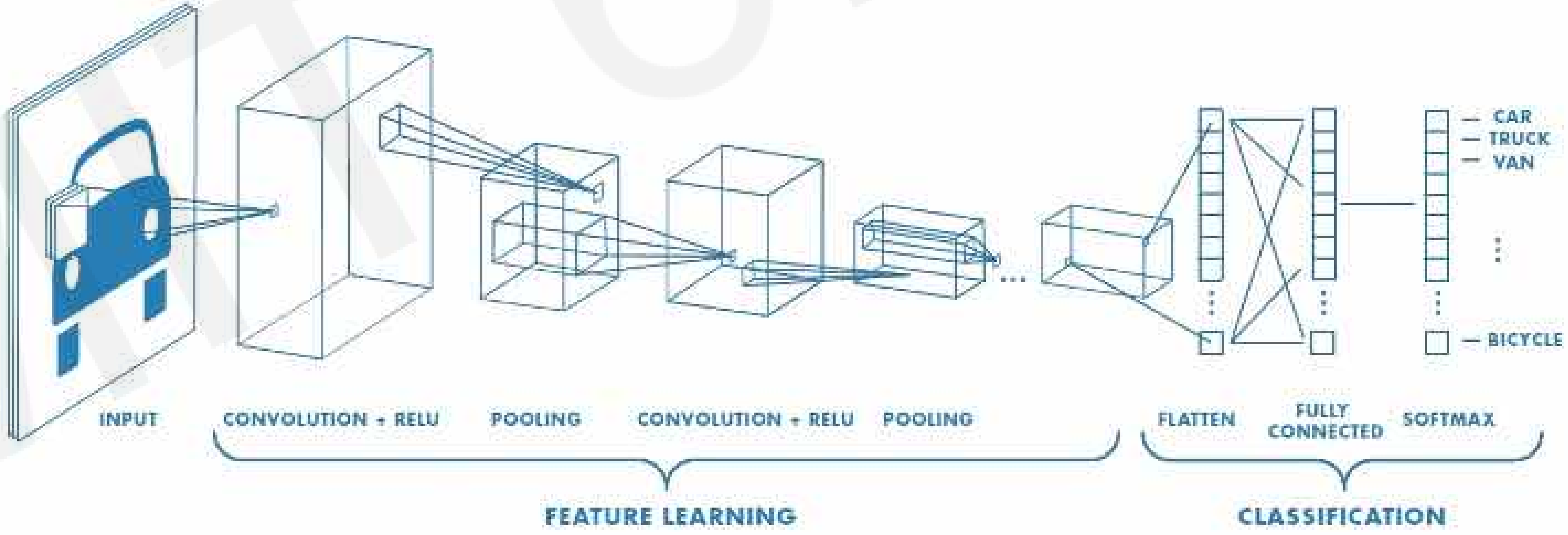
CNNs: Using Spatial Structure



1) Apply a set of weights to extract **local features**

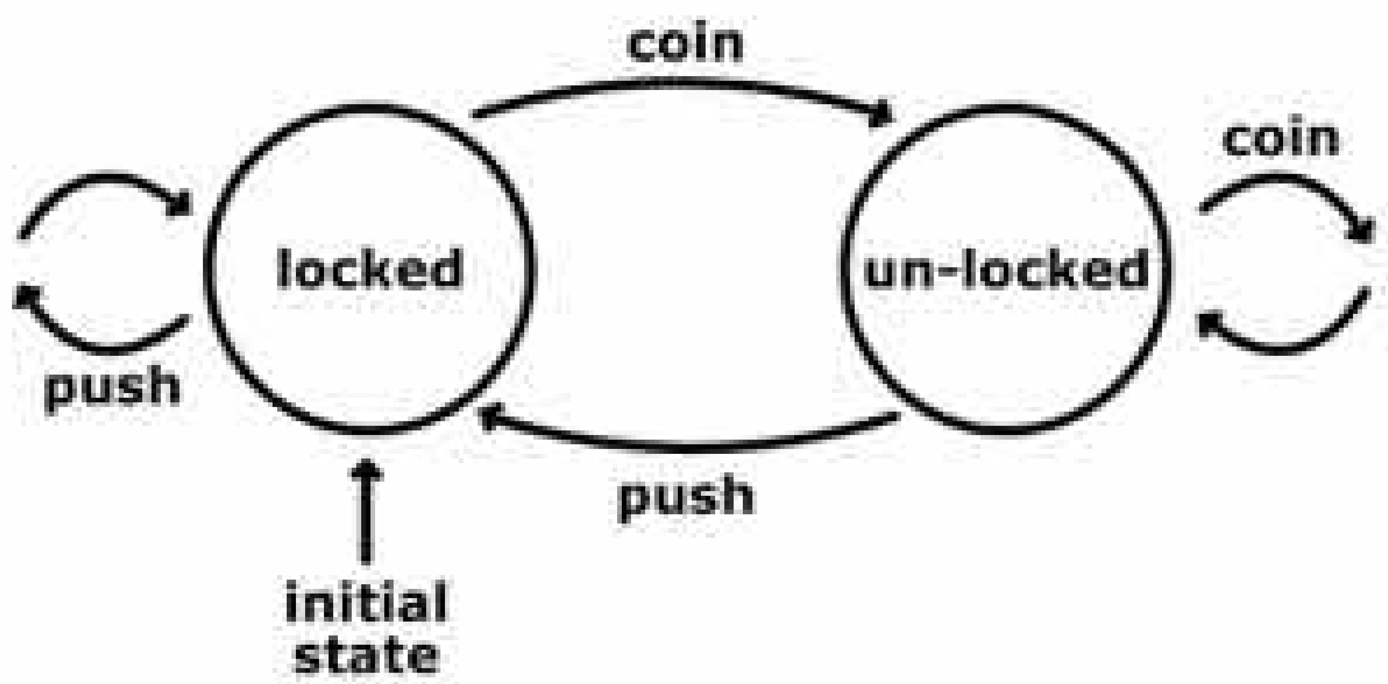
2) Use **multiple filters** to extract different features

3) **Spatially share** parameters of each filter

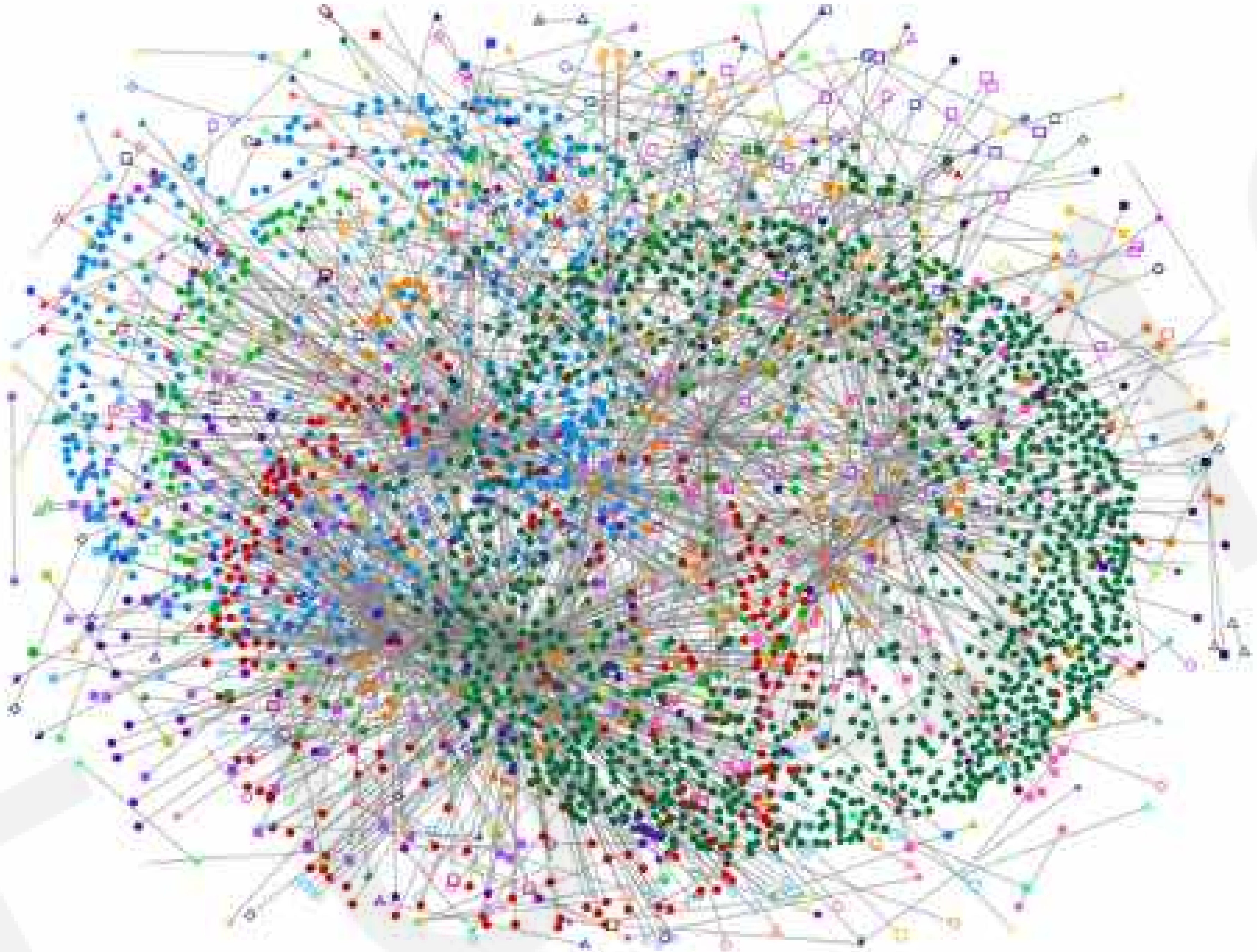


MIT 6.S191

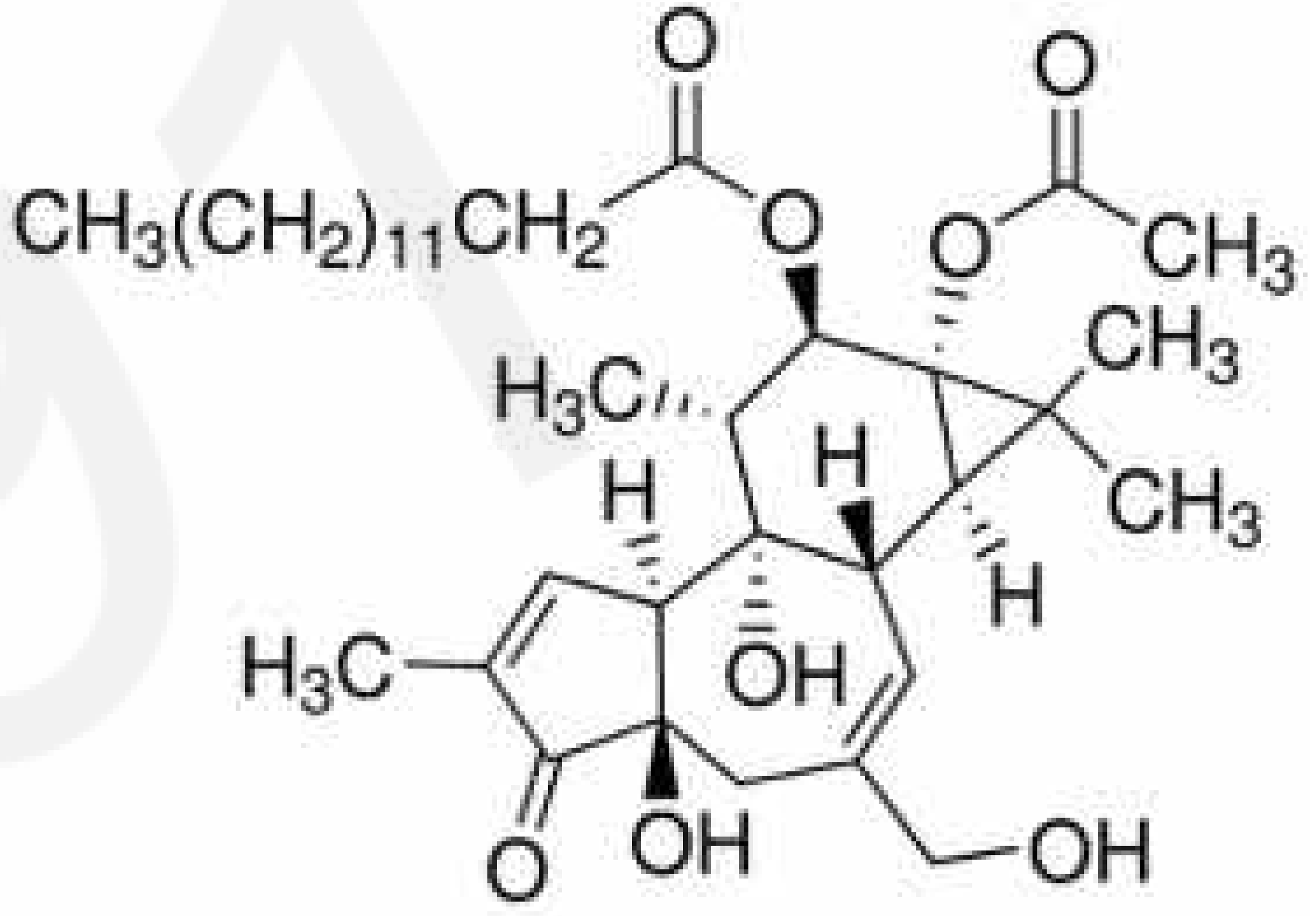
Graphs as a Structure for Representing Data



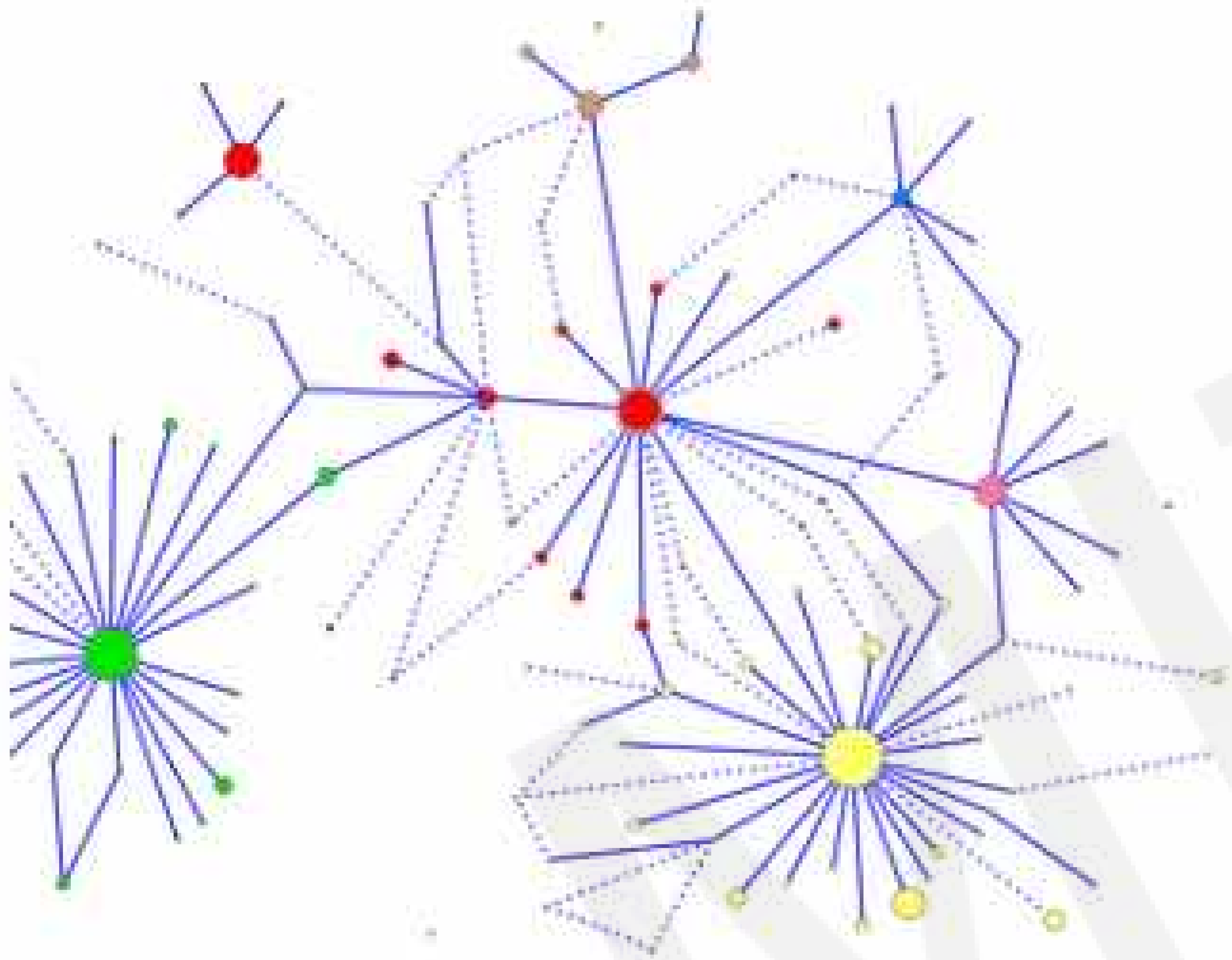
State Machines



Social Networks

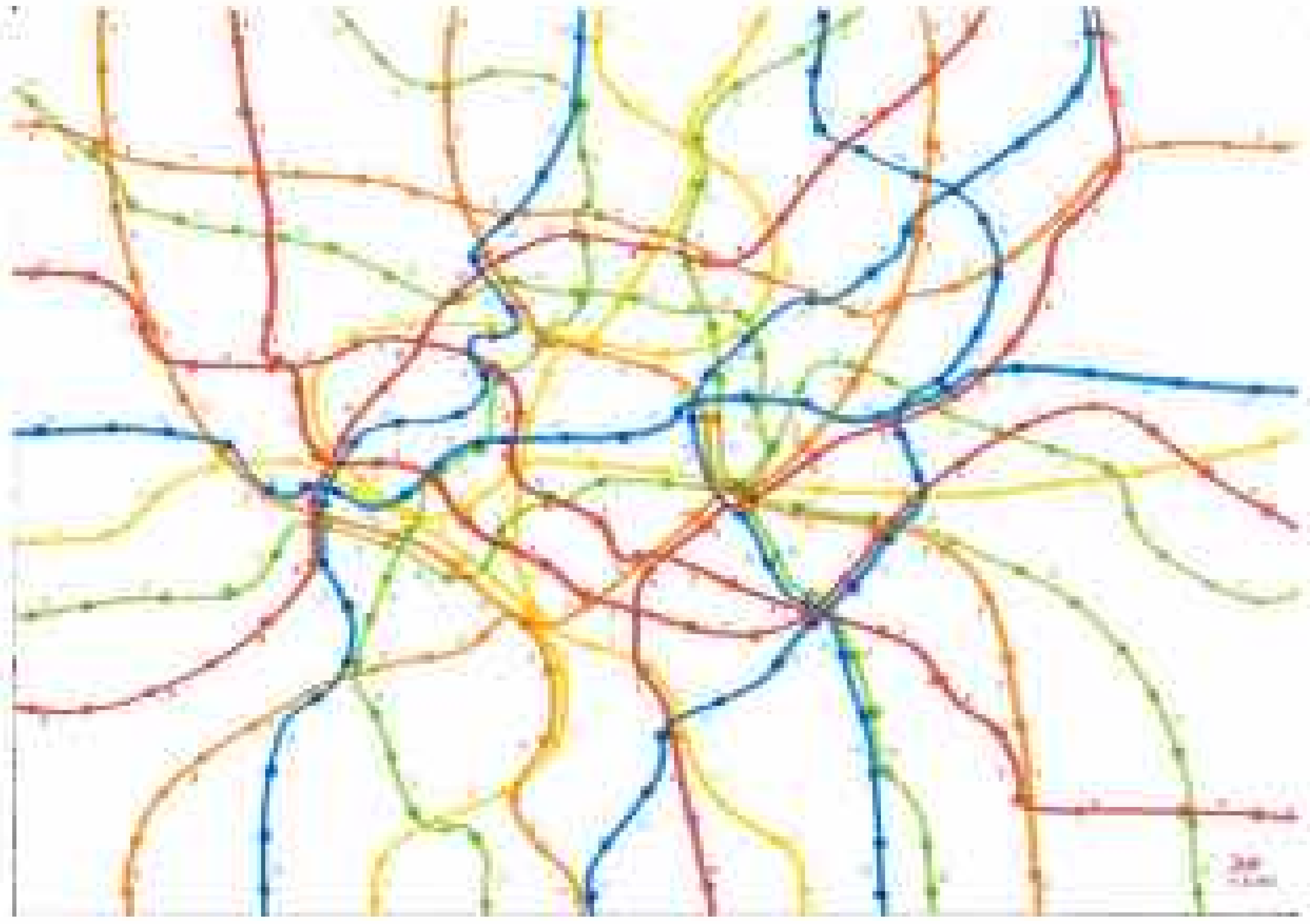


Molecules



Biological Networks

Many real-world data – such as networks – cannot be captured by "standard" encodings or Euclidean geometries



Mobility & Transport

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

Graph Convolutional Networks

Convolutional Networks

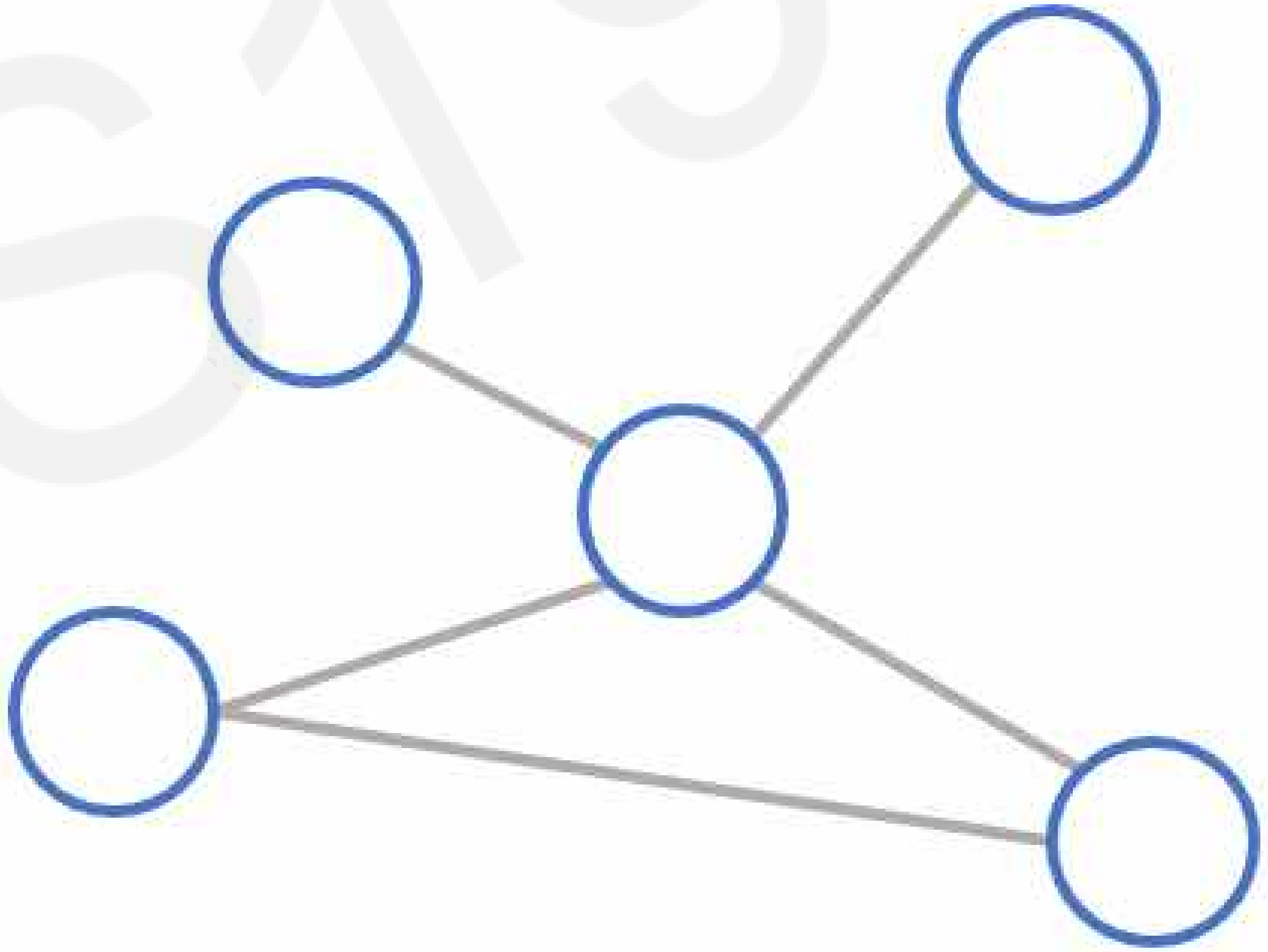
Graph Convolutional Networks (GCNs)

MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)



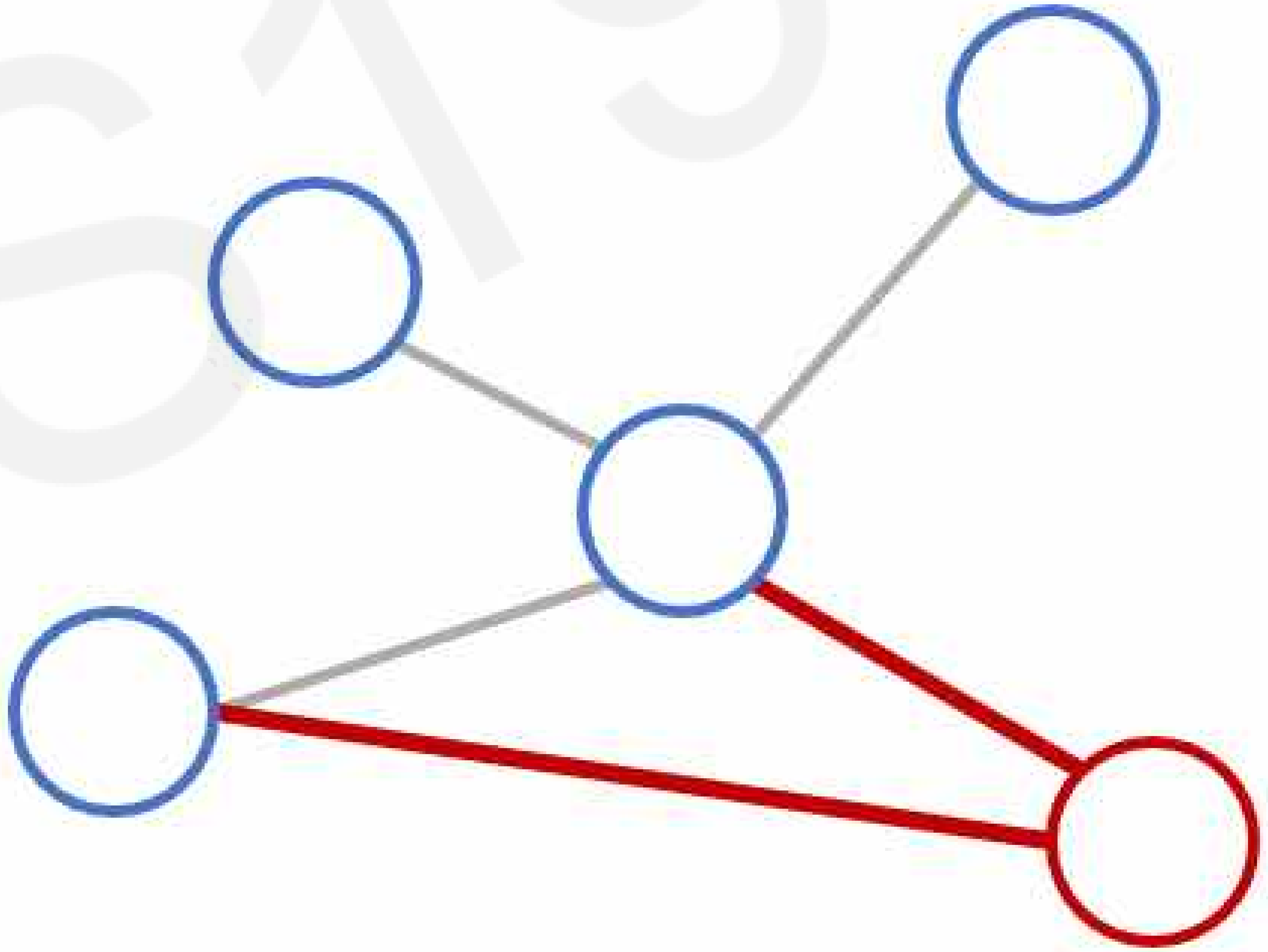
Duvenaud+ *NeurIPS* 2015; Kipf & Welling *ICLR* 2017.

MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

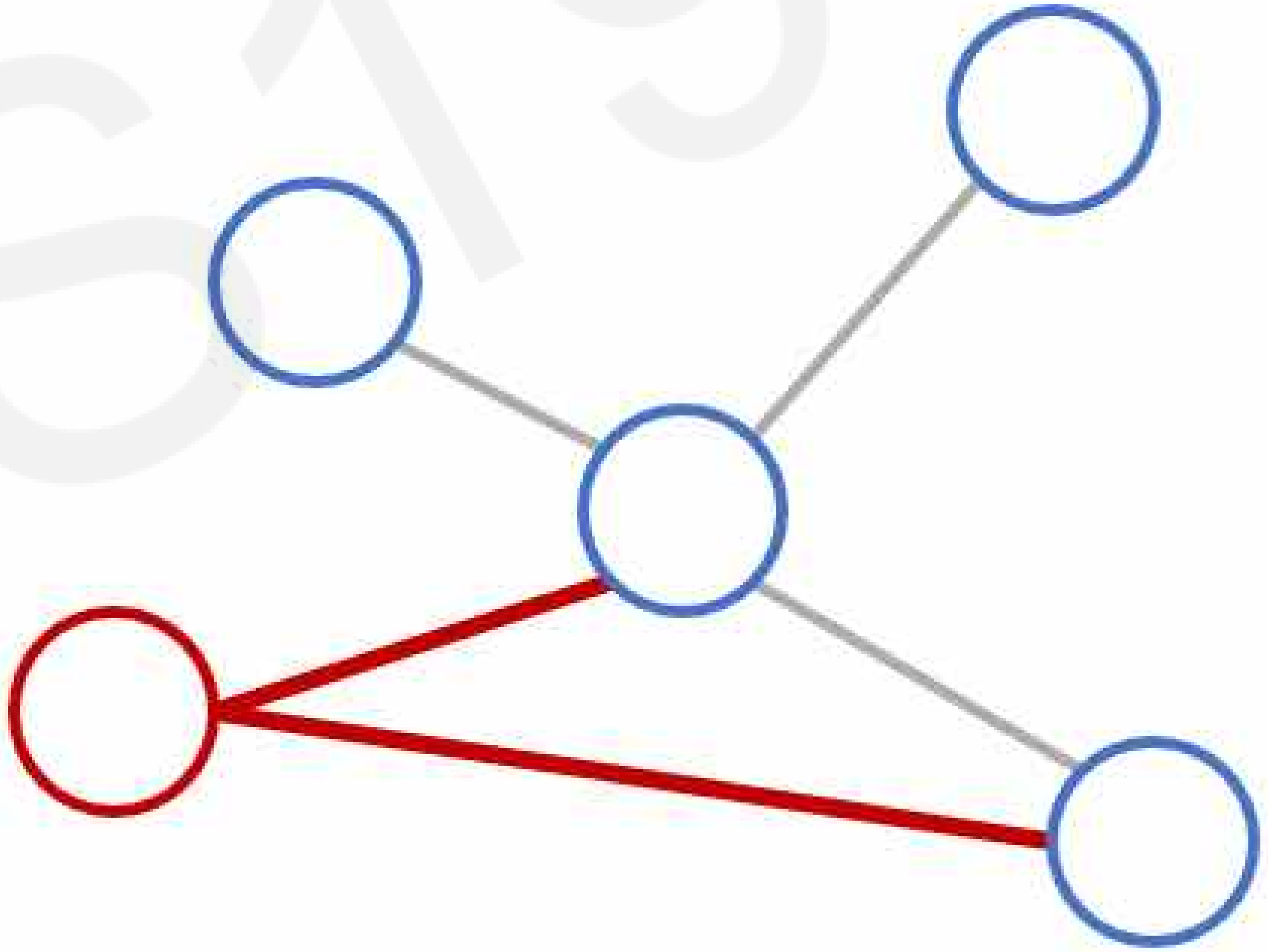


MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

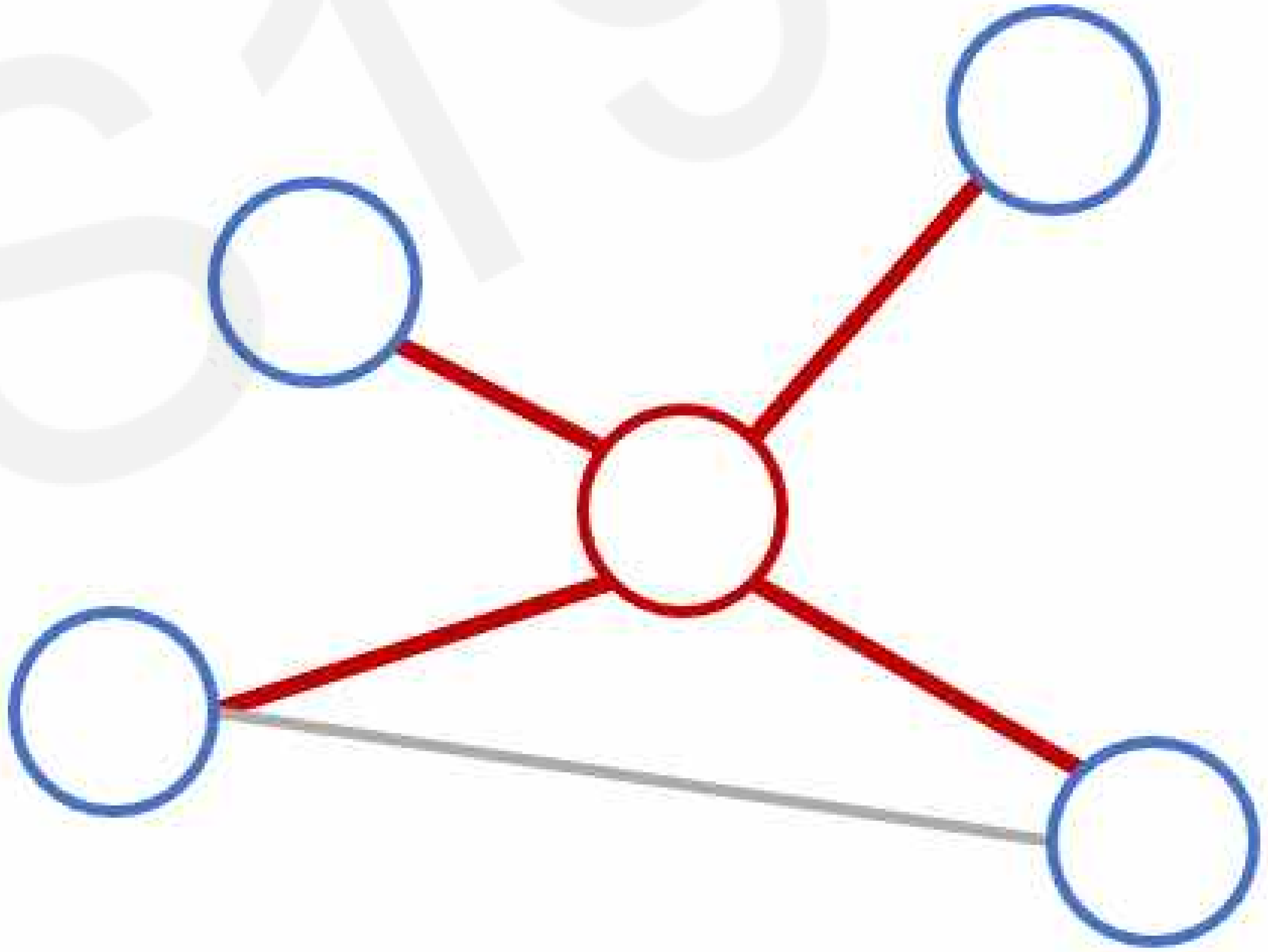
Graph Convolutional Networks (GCNs)



Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)



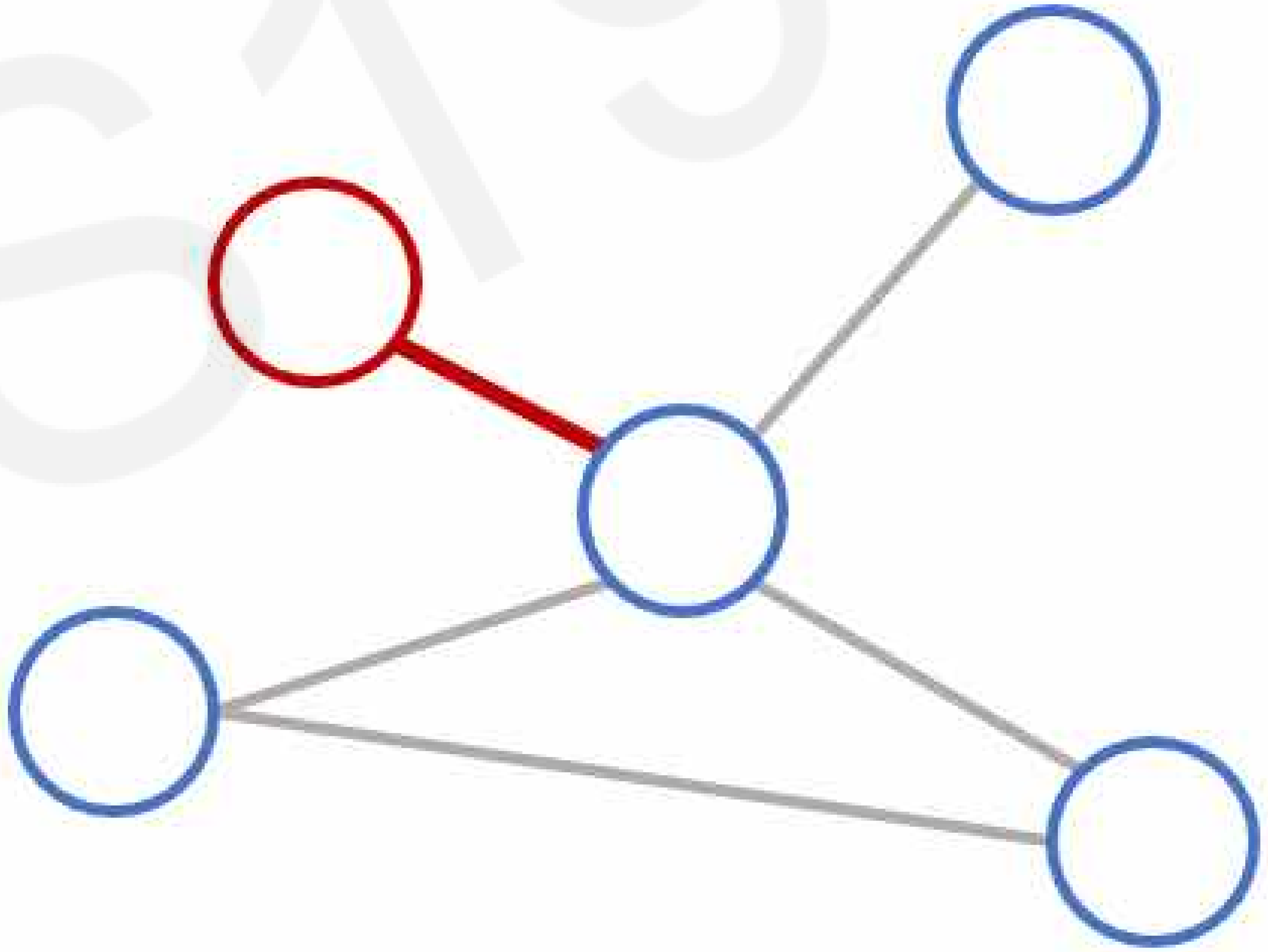
MIT 6.S191

Duvenaud+ *NeurIPS* 2015; Kipf & Welling *ICLR* 2017.

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

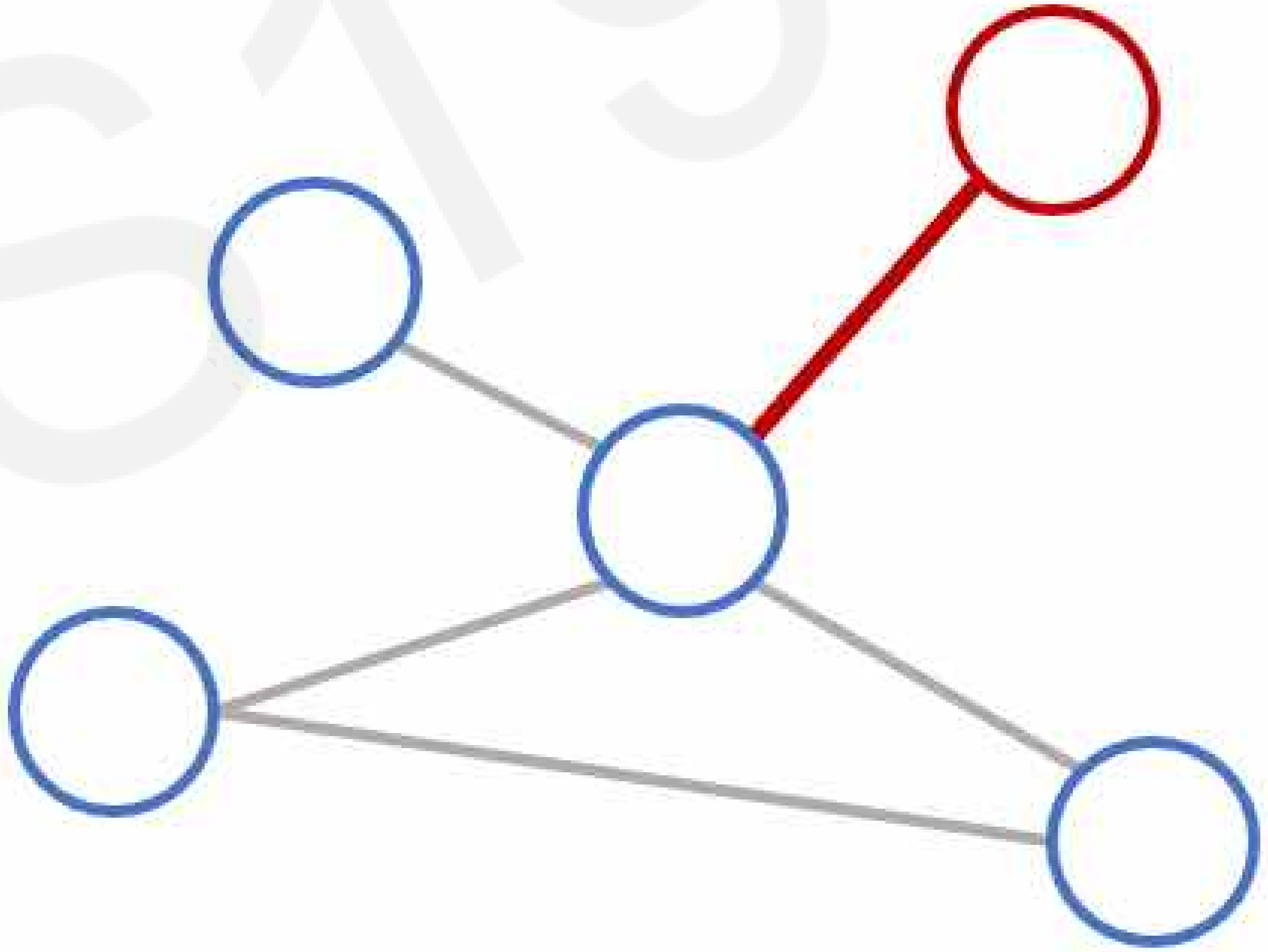


MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

Graph Convolutional Networks (GCNs)

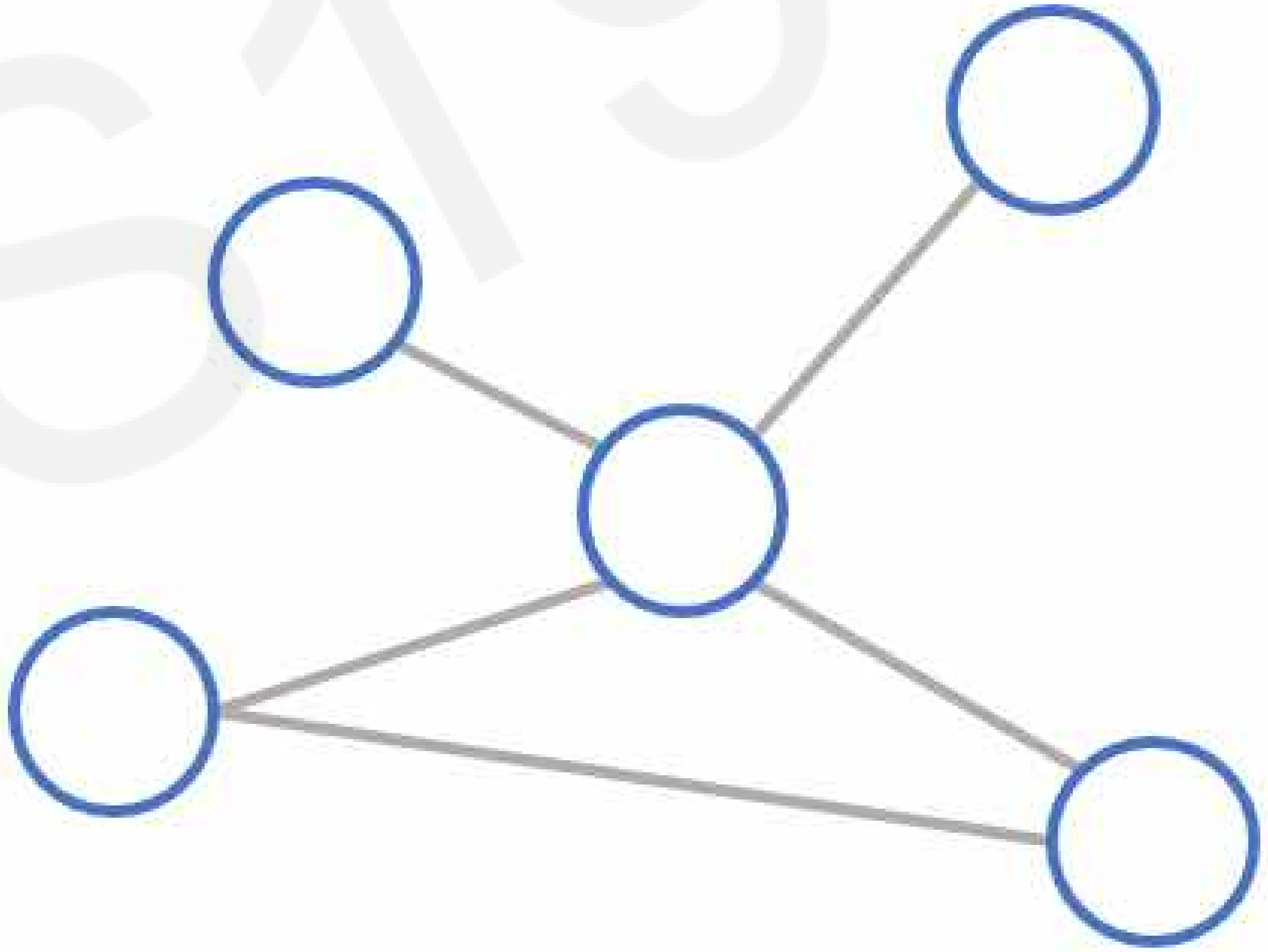


MIT 6.S191

Graph Convolutional Networks

Convolutional Networks

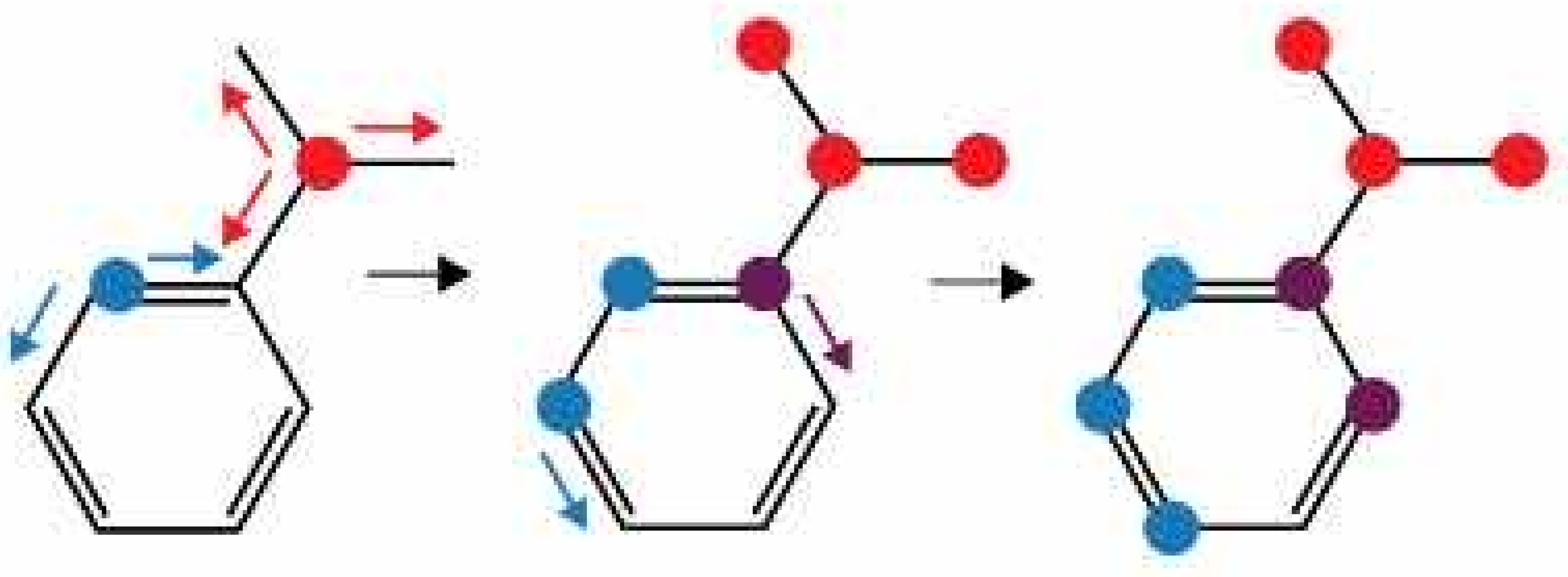
Graph Convolutional Networks (GCNs)



MIT 6.S191

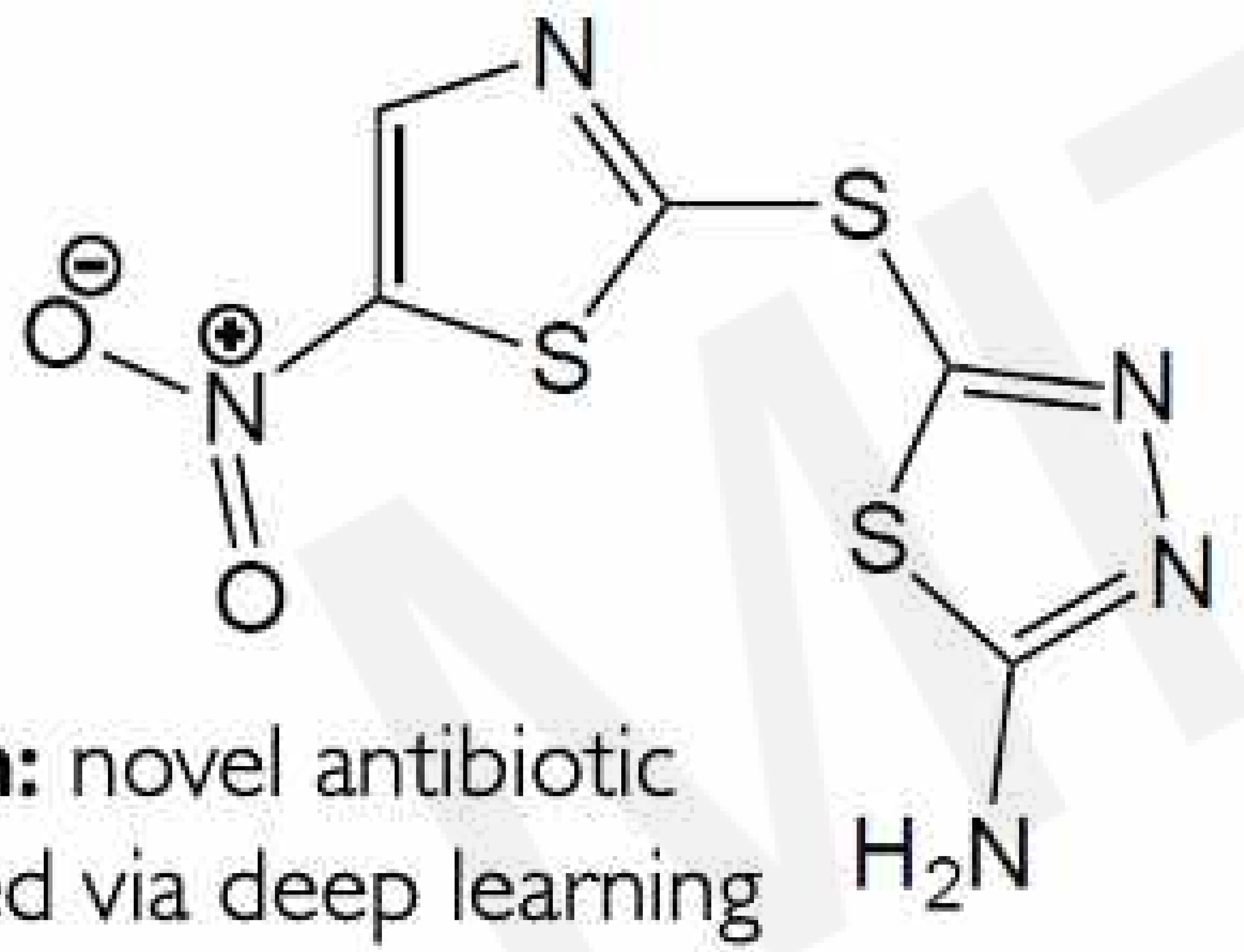
Applications of Graph Neural Networks

Molecular Discovery



Message-passing neural network

Jin+ *JCIM* 2019; Soleimany+ *ACS Cent. Sci.* 2021



Halicin: novel antibiotic discovered via deep learning

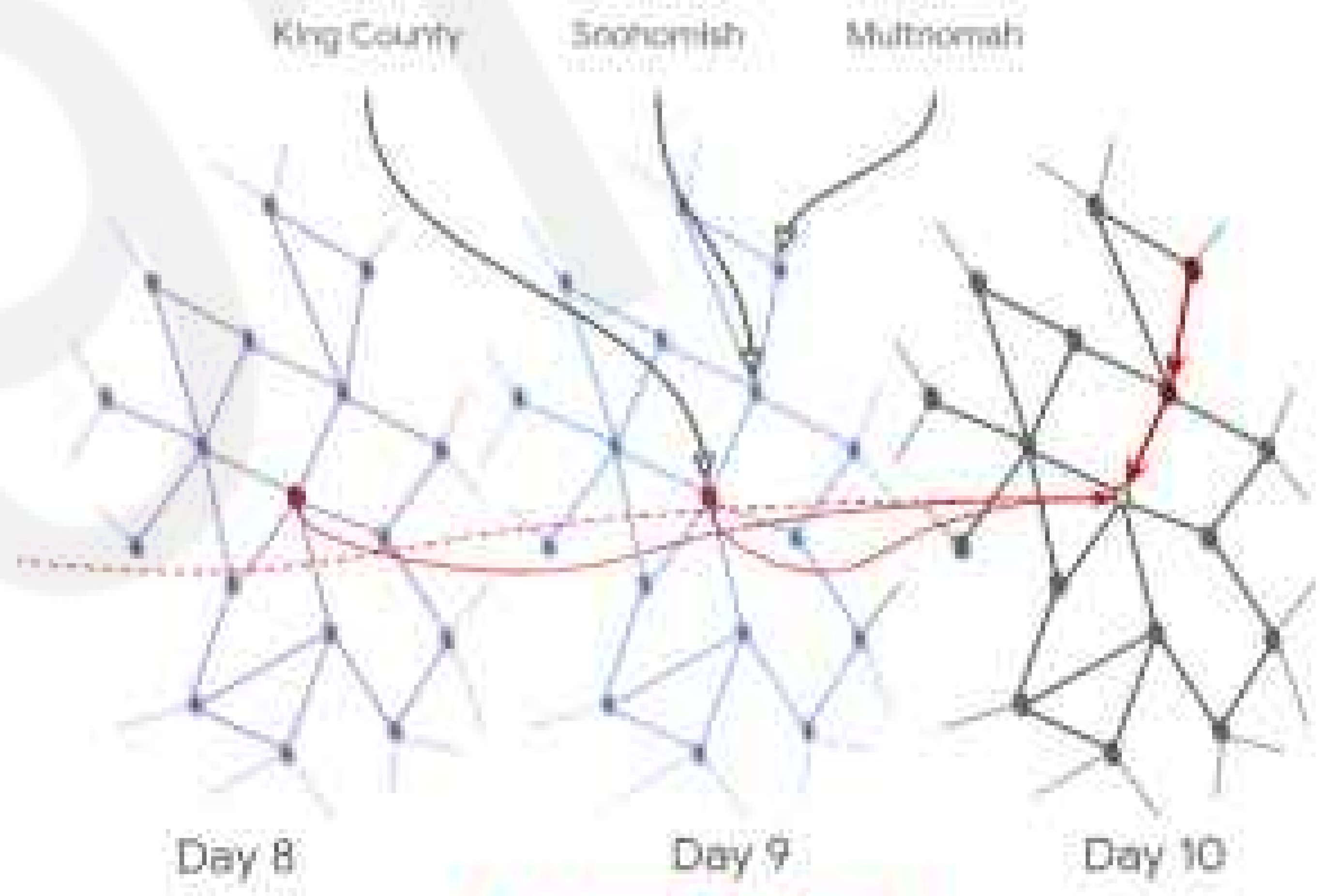
Stokes+ *Cell* 2020

Traffic Prediction

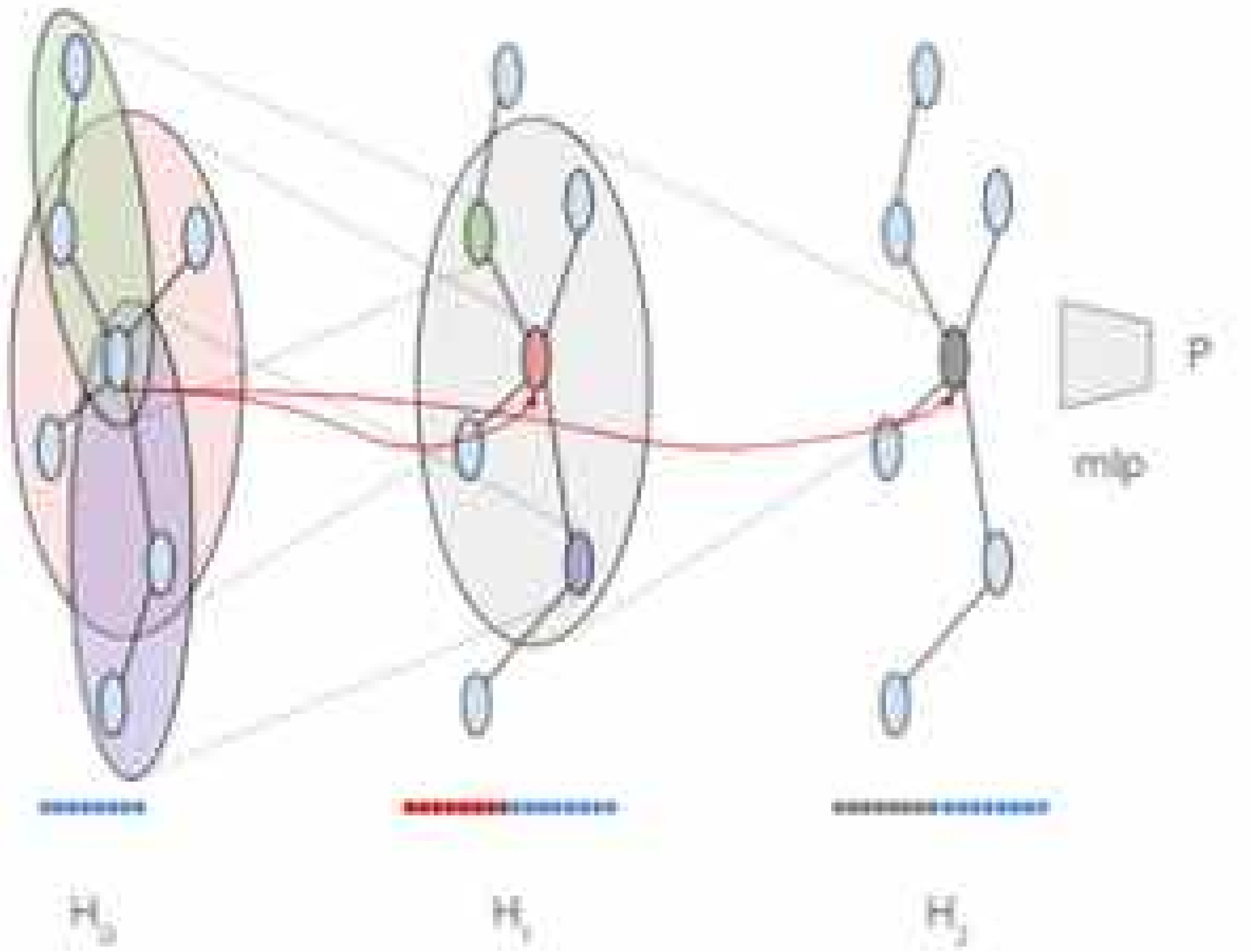
ETA Improvements with GoogleMaps

DeepMind + GoogleMaps

COVID-19 Forecasting



Spatio-temporal data



Graph network + temporal embedding

Kapoor+ *KDD* 2020

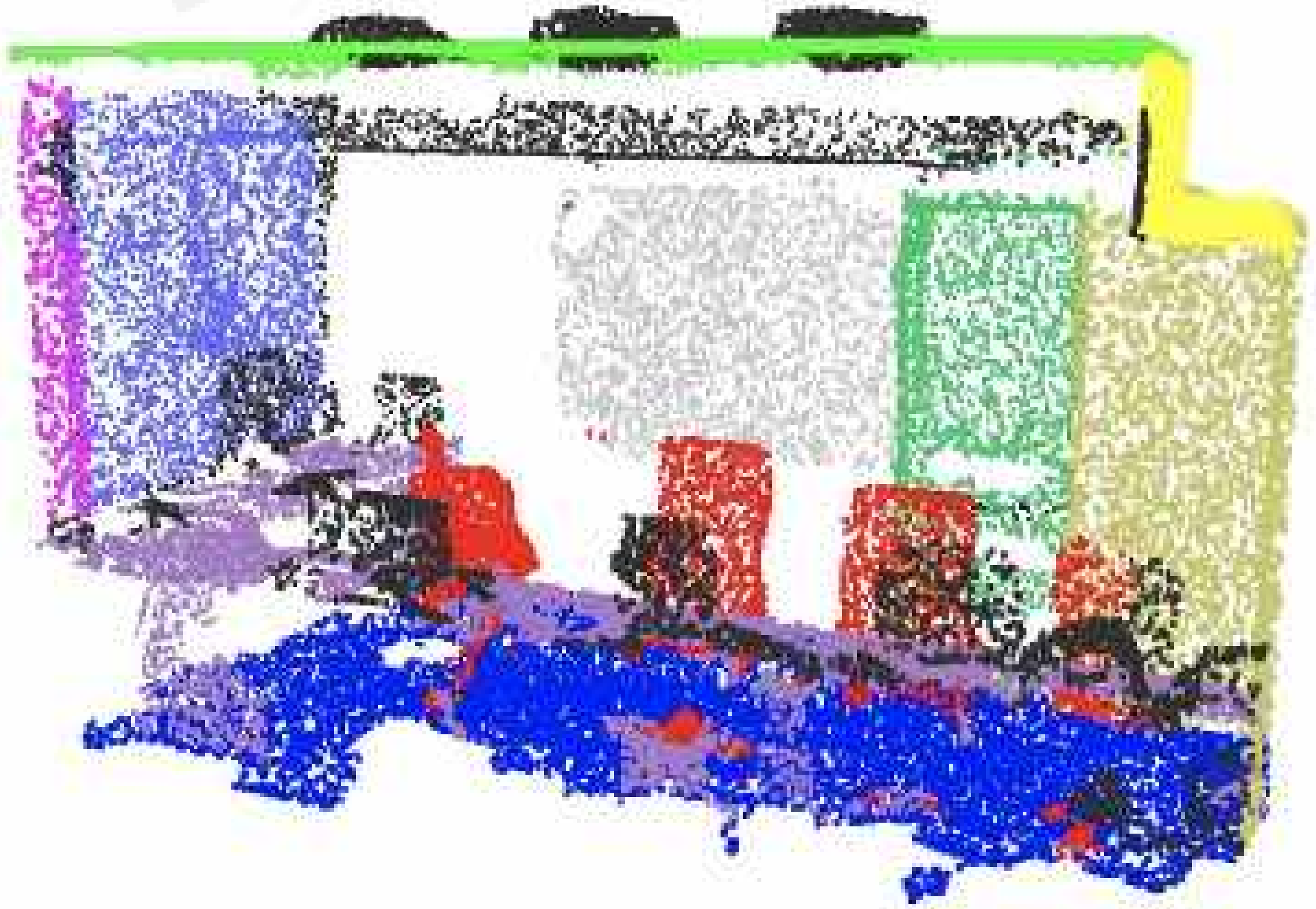
MIT 6.S191

Learning From 3D Data

Point clouds are **unordered sets** with **spatial dependence** between points

Classification

Part Segmentation

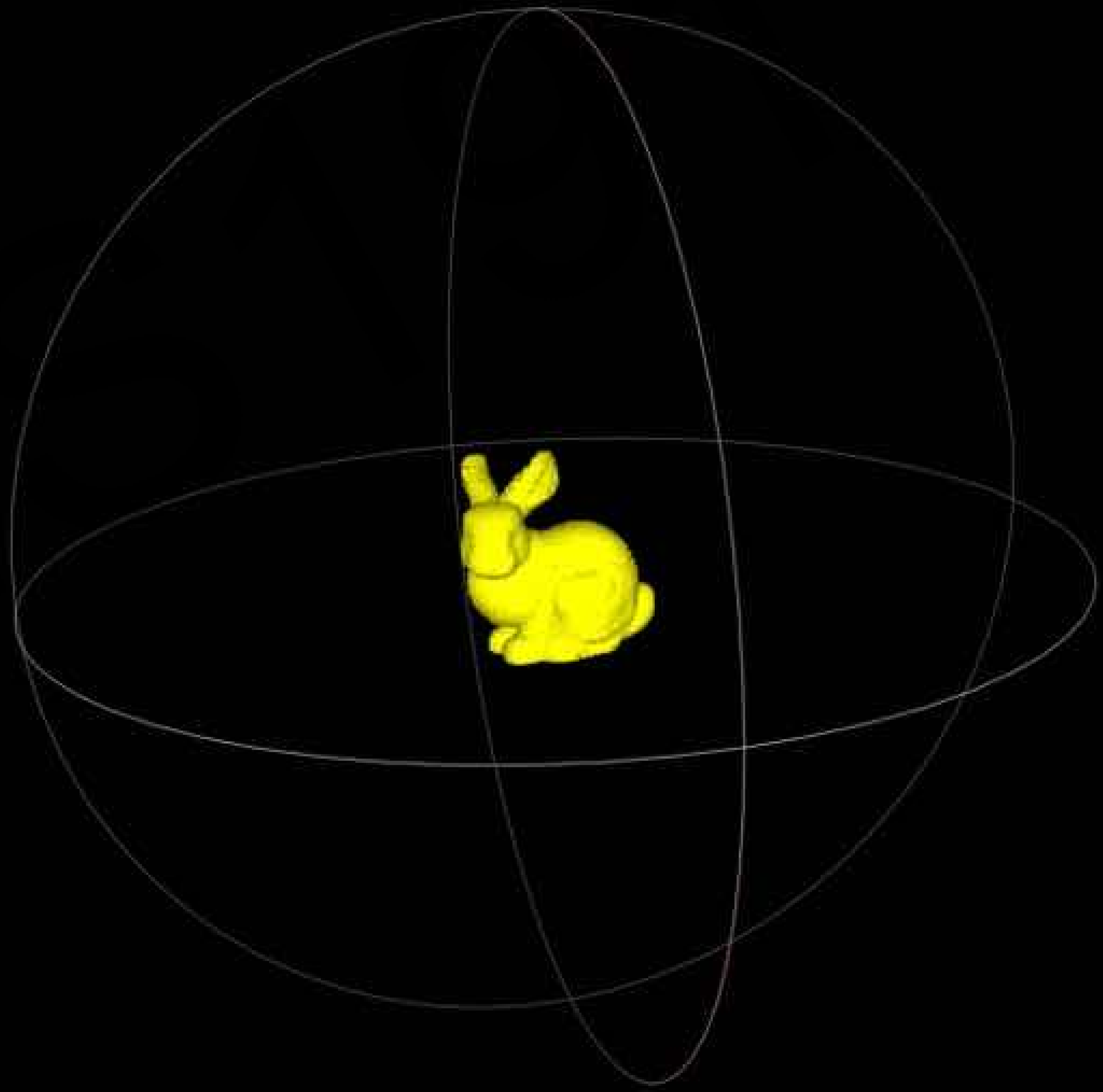


Semantic Segmentation

MIT 6.S191

Extending Graph CNNs to Pointclouds

Capture local geometric features of point clouds while maintaining order invariance

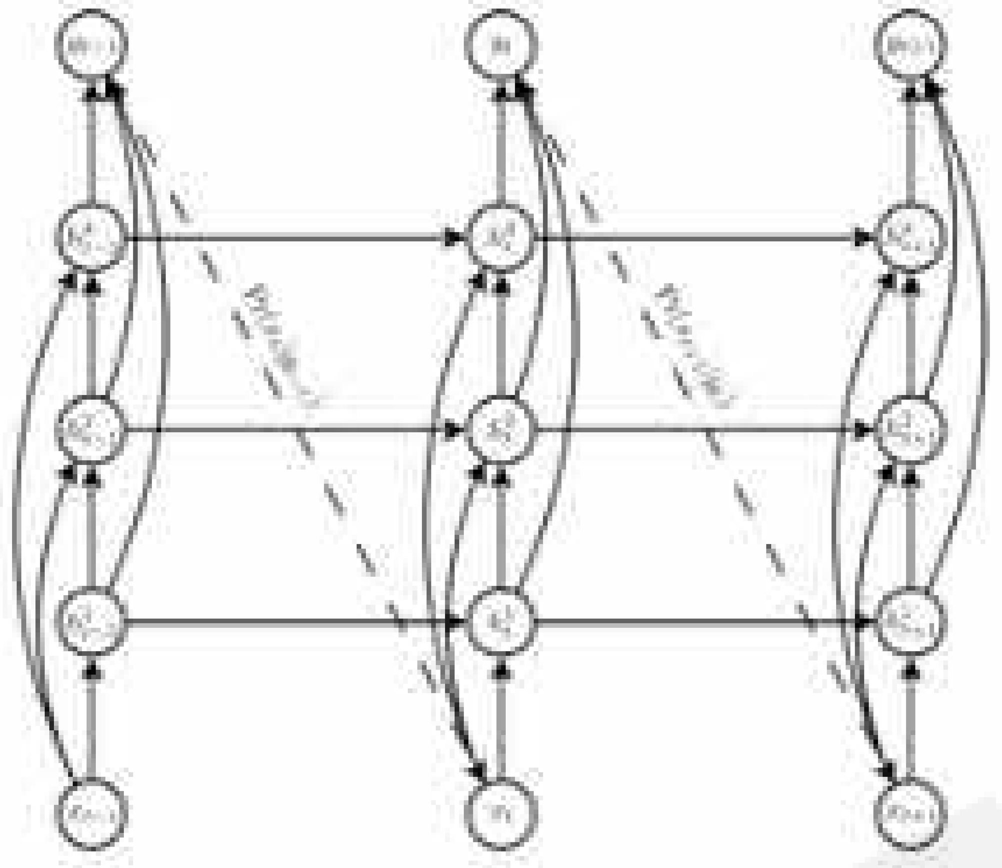


New Frontiers II:
Automated Machine Learning & AI

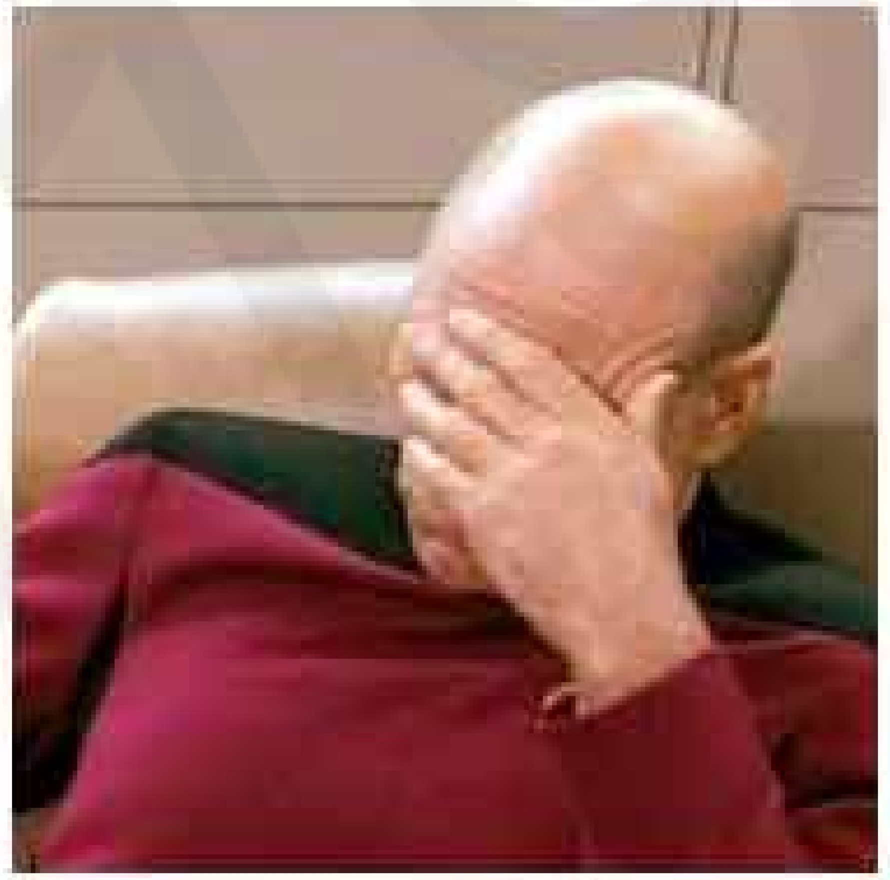
MIT 6.S191

Motivation: Automated Machine Learning

Standard deep neural networks are optimized for **a single task**



Complexity of models increases



Greater need for specialized engineers

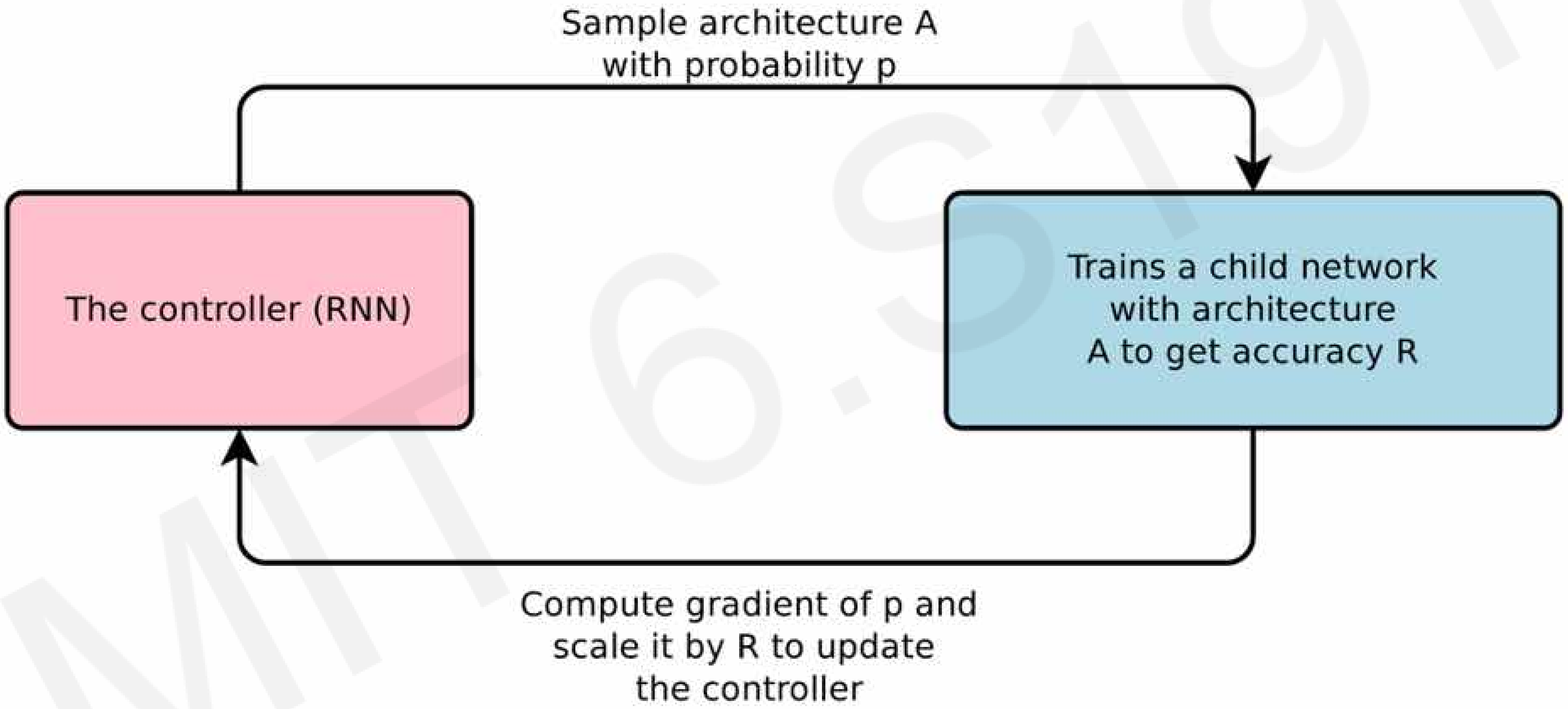
Often require **expert knowledge** to build an architecture for a given task

Build a learning algorithm that **learns which model** to use to solve a given problem

AutoML

MIT 6.S191

Automated Machine Learning (AutoML)



MIT 6.S191

Zoph and Le *ICLR* 2017.

AutoML: Model Controller

At each step, the model samples a brand new network

MIT 6.S191

AutoML: The Child Network

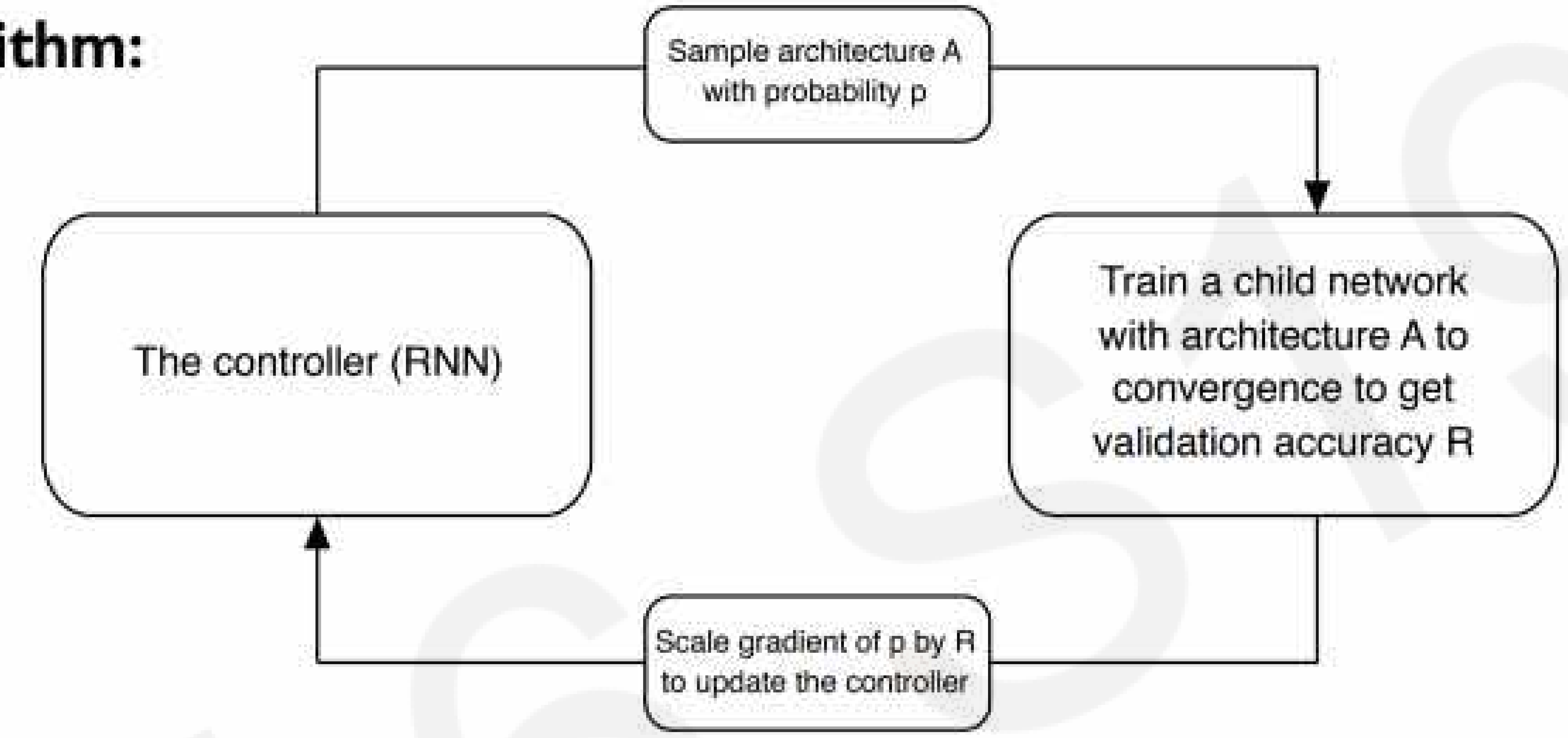
Training Data → Sampled network from RNN → Prediction

Compute final accuracy on this dataset.
Update RNN controller based on the accuracy of the child network after training.

MIT 6.S191

Learning Architectures for Image Recognition

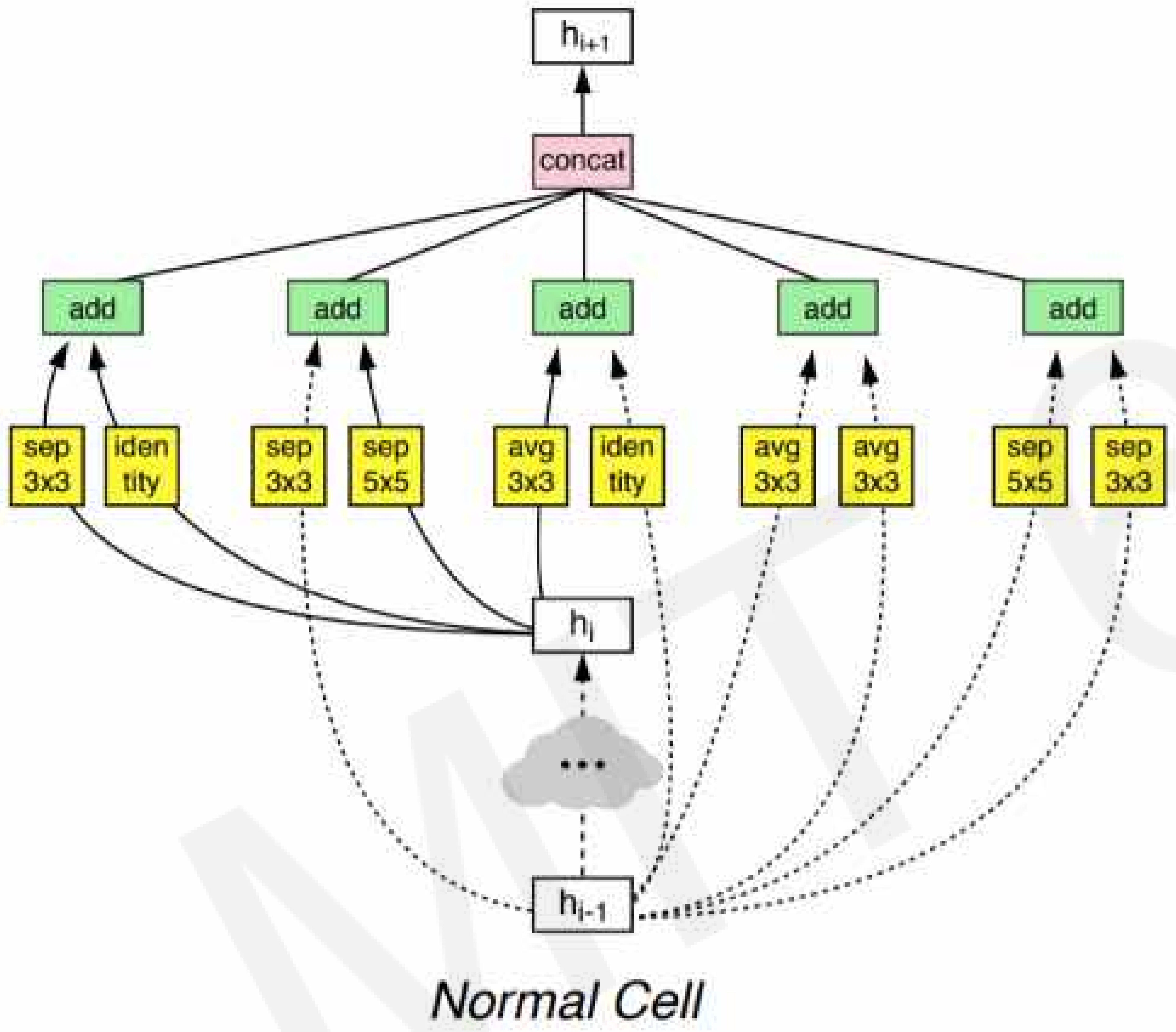
Neural architecture search algorithm:



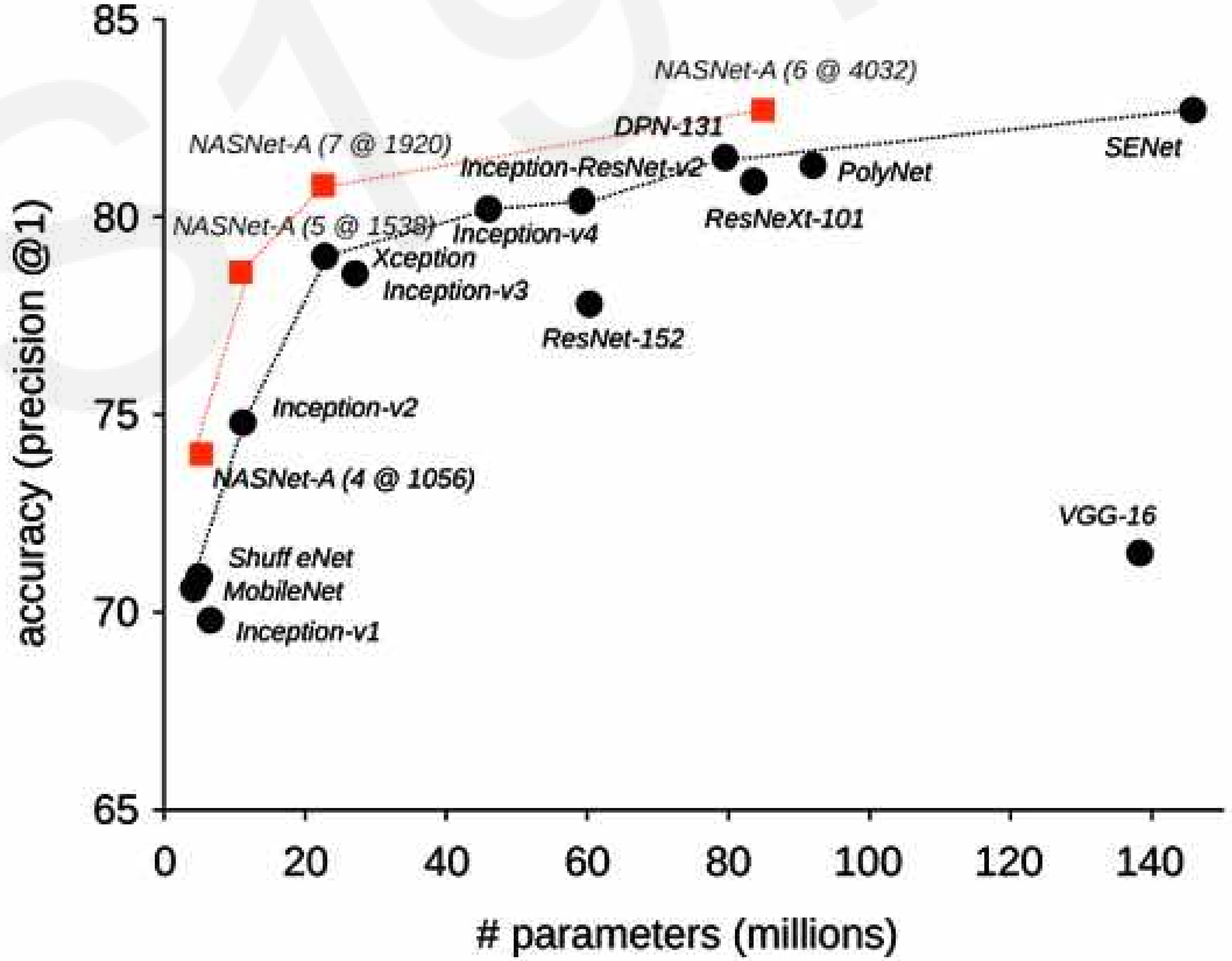
Controller architecture for constructing convolutional layers:

Learning Architectures for Image Recognition

Learned architecture for convolutional cell



Model performance on ImageNet



From AutoML to AutoAI

AutoAI

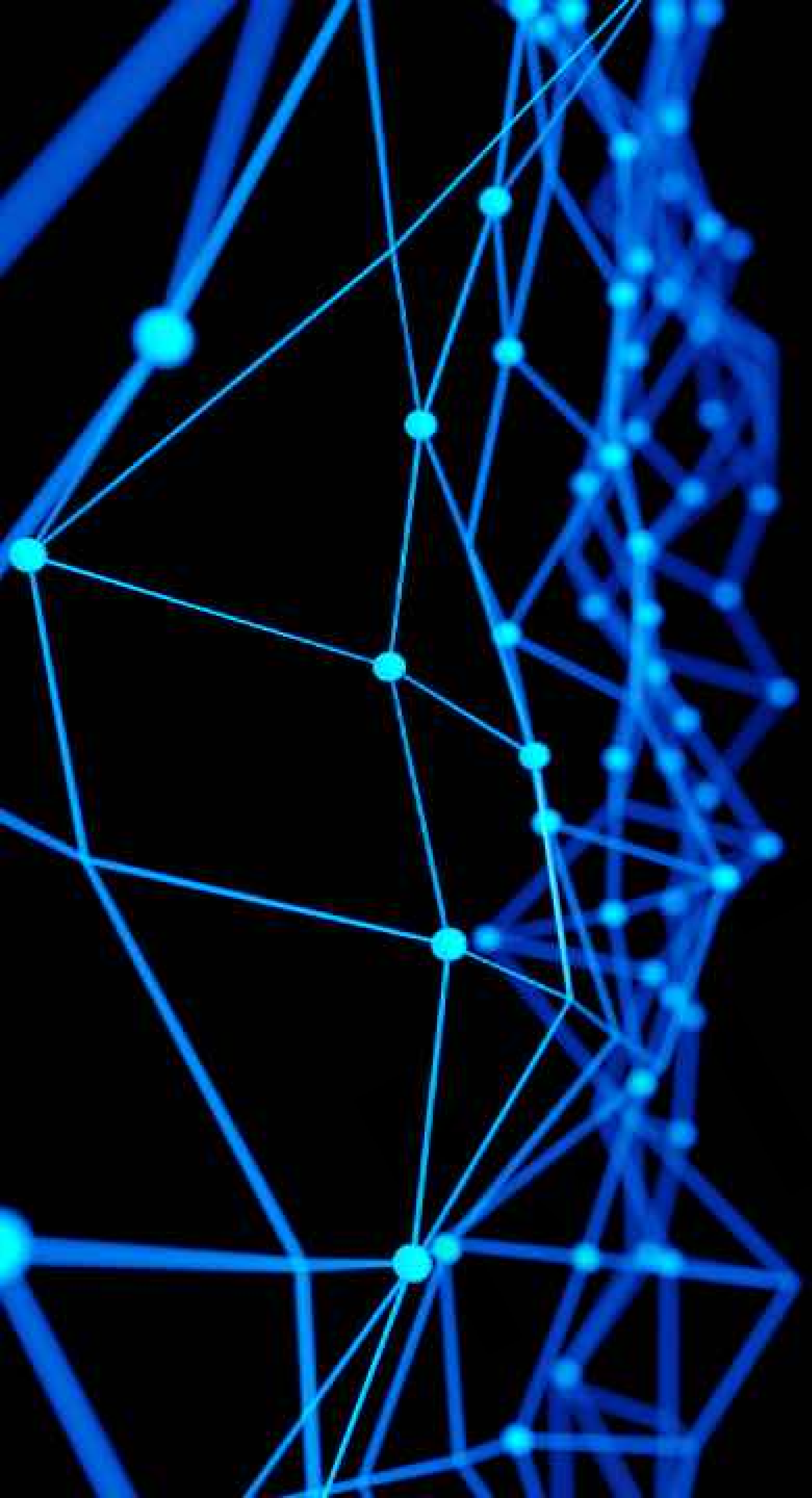
Provide data in a CSV file → Prepare data → Select model type → Generate and rank model pipelines → Save and deploy a model

Prepare data
- Feature type detection
- Missing values imputation
- Feature encoding and scaling

Select model type
- Selection of the best algorithm for the data

Generate and rank model pipelines
- Hyper-parameter optimization (HPO)
- Optimized feature engineering

MIT 6.S191



AutoAI Spawns a Powerful Idea

- Design an AI pipeline that can build new models capable of solving a task
- Reduces the need for experienced engineers to design the networks
- Makes deep learning more accessible to the public

Connections and distinctions between artificial and human intelligence

6.S191: Introduction to Deep Learning

Lab 3: Reinforcement Learning

Link to download labs:
http://introtodeeplearning.com#schedule

1. Open the lab in Google Colab
2. Start executing code blocks and filling in the #TODOs
3. Need help? Come to 10-250/Gather.Town!