

KNOWLEDGE, LEARNING & INFERENCE

Paula Parpart

Dept. of Experimental Psychology

09.01.2016, KLI Lecture

Computational Models of Cognition:

- I. Bayesian Models of Cognition**
- II. The Future of AI: Human vs Machine Learning**

Paula Parpart

Experimental Psychology

09.01.2016, KLI Lecture

COURSE OVERVIEW

- 16/01/2016 – Paula Parpart: Computational Models of Cognition
- 23/01/2016 – Maarten Speekenbrink: Instance vs. Abstraction Learning
- 30/01/2016 – Kurt Braunlich: Category Learning and the Brain
- 06/02/2016 – Eric Schulz: Reinforcement & Function Learning
- tbc - Dave Lagnado: Bayesian Networks in Evidential Reasoning

--- READING WEEK ---

- 20/02/2016 – Christos Bechlivanidis: Causal Models: Representation, Learning and Inference
- 27/02/2016 – Paula Parpart: Active Learning
- 06/03/2016 – Paula Parpart: Heuristics as Bayesian Inference
- 13/03/2016 – Adam Harris: Bayesian Argumentation
- 20/03/2016 – Brad Love: A cognitive science informed Turing Test



Personal
research

COURSE OVERVIEW



Learning

- 16/01/2016 – Paula Parpart: Computational Models of Cognition
- 23/01/2016 – Maarten Speekenbrink: Instance vs. Abstraction **Learning**
- 30/01/2016 – Kurt Braunlich: Category **Learning** and the Brain
- 06/02/2016 – Eric Schulz: Reinforcement & Function **Learning**
- tbc - Dave Lagnado: Bayesian Networks in Evidential Reasoning

--- READING WEEK ---

- 20/02/2016 – Christos Bechlivanidis: Causal Models: Representation, **Learning** and Inference
- 27/02/2016 – Paula Parpart: Active **Learning**
- 06/03/2016 – Paula Parpart: Heuristics as Bayesian Inference
- 13/03/2016 – Adam Harris: Bayesian Argumentation
- 20/03/2016 – Brad Love: A cognitive science informed Turing Test (**Learning**)

COURSE OVERVIEW



Representation
(= Knowledge)

- 16/01/2016 – Paula Parpart: Computational Models
- 23/01/2016 – Maarten Speekenbrink: Instance vs. Abstract
- 30/01/2016 – Kurt Braunlich: Category Learning and the Brain
- 06/02/2016 – Eric Schulz: Reinforcement & Function Learning
- tbc - Dave Lagnado: Bayesian Networks in Evidential Reasoning

(Representation & Inference)

--- READING WEEK ---

- 20/02/2016 – Christos Bechlivanidis: Causal Models: **Representation**, Learning and Inference
- 27/02/2016 – Paula Parpart: Active Learning (**Representation & Learning**)
- 06/03/2016 – Paula Parpart: Heuristics as Bayesian Inference
- 13/03/2016 – Adam Harris: Bayesian Argumentation
- 20/03/2016 – Brad Love: A cognitive science informed Turing Test (**Representation**)

COURSE OVERVIEW



Inference

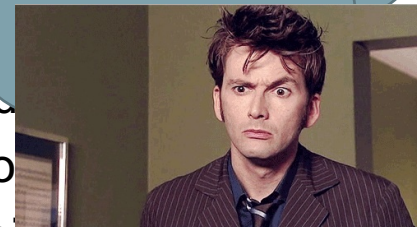
- 16/01/2016 – Paula Parpart: **Computational Models**
- 23/01/2016 – Maarten Speekenbrink: Instance vs. Abstra
- 30/01/2016 – Kurt Braunlich: Category Learning and the Brain
- 06/02/2016 – Eric Schulz: Reinforcement & Function Learning
- tbc - Dave Lagnado: Bayesian Networks in Evidential Reasoning
(Representation & **Inference**)

--- READING WEEK ---

- 20/02/2016 – Christos Bechlivanidis: Causal Models: Representation, Learning and **Inference**
- 27/02/2016 – Paula Parpart: Active Learning
- 06/03/2016 – Paula Parpart: Heuristics as Bayesian **Inference**
- 13/03/2016 – Adam Harris: Bayesian Argumentation (**Inference**)
- 20/03/2016 – Brad Love: A cognitive science informed Turing Test (**Inference**)

COURSE OVERVIEW

Bayesian?



- 16/01/2016 – Paula Parpart: **(Bayesian)** Computational Models
- 23/01/2016 – Maarten Speekenbrink: Instance vs. Abstraction
- 30/01/2016 – Kurt Braunlich: Category Learning and the Brain
- 06/02/2016 – Eric Schulz: Reinforcement & Function Learning
- tbc - Dave Lagnado: **Bayesian** Networks in Evidential Reasoning

--- READING WEEK ---

- 20/02/2016 – Christos Bechlivanidis: Causal Models: Representation, Learning and Inference
- 27/02/2016 – Paula Parpart: Active Learning
- 06/03/2016 – Paula Parpart: Heuristics as **Bayesian** Inference
- 13/03/2016 – Adam Harris: **Bayesian** Argumentation (Inference)
- 20/03/2016 – Brad Love: A cognitive science informed Turing Test (Inference)

Logistics

- Readings for all lectures are uploaded on Moodle.
 - Read at least one before the lecture
- Lecture slides will be on Moodle before or slightly after each lecture.
- Lectures are usually recorded – recordings will appear on moodle
- Assessment: 100% coursework
 - Essay topics will be announced on Moodle
 - Deadline: 27/04/2017
- For any lecture related question email the lecturer
- For any general issue about the module email Paula (paula.parpert@ucl.ac.uk)



MORE DISCUSSION!

TOPIC OVERVIEW

1. Bayesian Models of Cognition

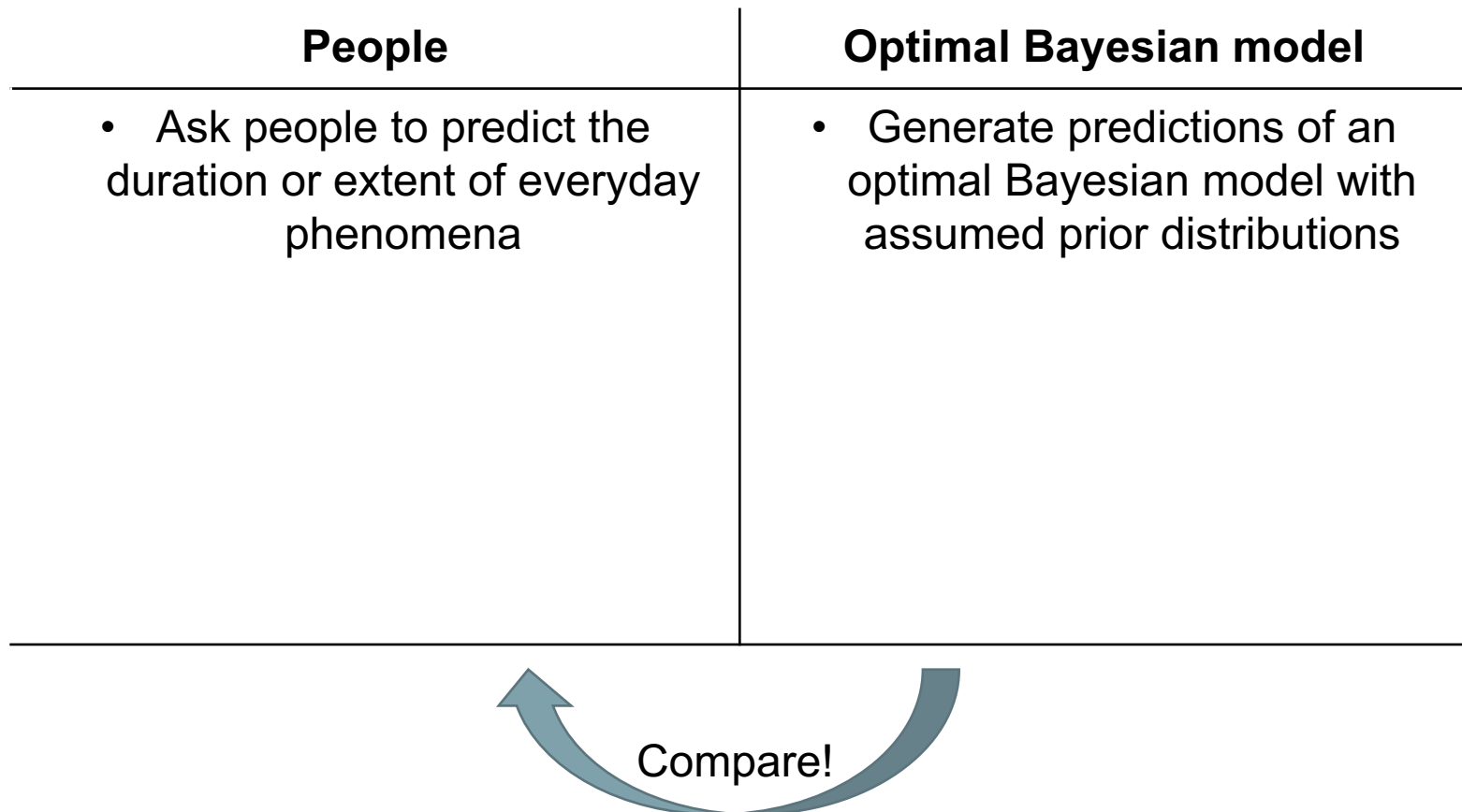
If you were assessing the prospects of a 60-year-old man, how much longer would you expect him to live?

- Life poses many of these inductive problems where the true answer cannot be determined on the basis of the limited data available
- Yet common sense suggests at least reasonable guess
- → Accounts of perception and memory suggest that these systems effectively approximate optimal statistical inference, correctly **combining new data with an accurate probabilistic model of the environment** (Anderson, 1990; Anderson & Schooler, 1991)

Contrasting hypotheses

traditionally	Griffiths & Tenenbaum (2006)
<ul style="list-style-type: none"><li data-bbox="185 639 919 853">• Cognitive everyday judgements are error prone due to the use of heuristics (Kahneman & Tversky's work)<li data-bbox="227 982 877 1082">• Heuristics are insensitive to prior probabilities	<ul style="list-style-type: none"><li data-bbox="996 639 1673 911">• People are near-optimal and are able to make smart predictions from sparse data even with everyday cognitive judgements<li data-bbox="1025 982 1702 1139">• Near-optimal -> optimal statistical inference = optimal Bayesian inference

How do cognitive judgements compare with optimal statistical inference?



Bayes Theorem



- Task: Predicting total life span of a man we just met, on the basis of the man's current age.

$$p(t_{total}|t) \propto p(t|t_{total})p(t_{total})$$

- t_{total} = total amount of time the man will live
- t = current age

Bayes Theorem

- Task: Predicting total life span of a man we just met, on the basis of the man's current age.

$$p(t_{total}|t) \propto \underbrace{p(t|t_{total})}_{\text{likelihood}} p(t_{total})$$

- $p(t|t_{total})$ = likelihood is the probability of first encountering a man at age t given that his total life span is t_{total}
- For simplicity, we assume we are equally likely to meet a man at any point in his life, so probability is uniform $p(t|t_{total}) = 1/t_{total}$.

Bayes Theorem

- Task: Predicting total life span of a man we just met, on the basis of the man's current age.

$$p(t_{total}|t) \propto p(t|t_{total}) \underbrace{p(t_{total})}_{\text{prior}}$$

- $p(t_{total})$ = prior probability reflects our general expectations about the relevant class of events – about how likely it is that a man's life span will be t_{total} .
- Actuarial data shows that the distribution of life spans in our society is approximately Gaussian – normally distributed- with mean μ of 75 years and st.dev. σ of 16 years.

Bayes Theorem

- Task: Predicting total life span of a man we just met, on the basis of the man's current age.

$$\underbrace{p(t_{total}|t)}_{\text{posterior}} \propto \underbrace{p(t|t_{total})}_{\text{likelihood}} \underbrace{p(t_{total})}_{\text{prior}}$$

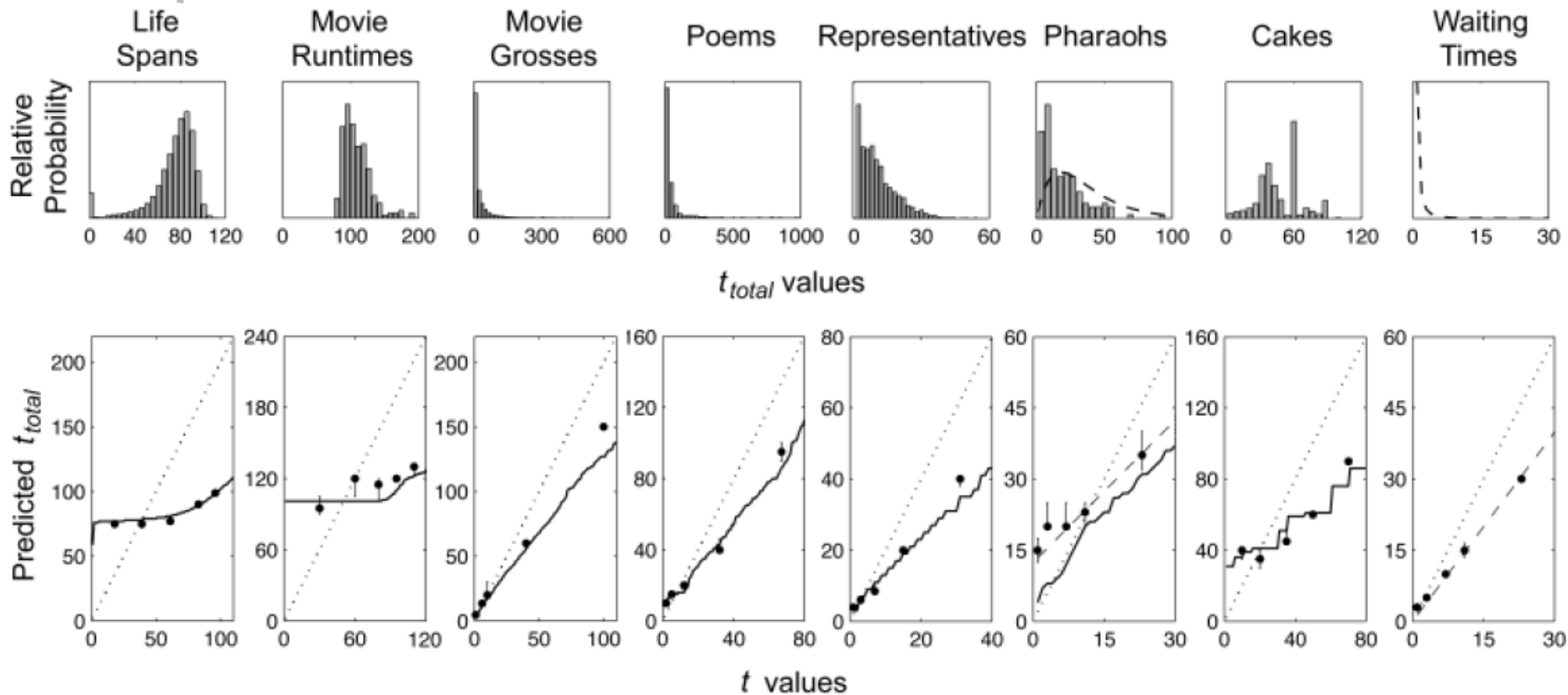
my updated beliefs (pointing to the posterior term)
 prior beliefs (pointing to the prior term)

- Combining the prior with the likelihood yields a probability distribution $p(t_{total}|t)$ over all possible total life spans t_{total} for a man encountered at age t
- A good guess for t_{total} is the median of this distribution (point at which it is equally likely that true life span is longer or shorter)

Bayesian prediction function

Different priors for different phenomena

Empirical prior distributions



If you were the executive evaluating the performance of a movie that had made \$40 million at the box office so far, what would you estimate for its total sales?

Results

- People's judgements for life spans, movie run times, etc. **were indistinguishable from optimal Bayesian predictions** based on the empirical prior distributions.

traditionally	Griffiths & Tenenbaum (2006)
<ul style="list-style-type: none"> • Cognitive everyday judgements are error prone due to the use of heuristics (Kahneman & Tversky's work) 	<ul style="list-style-type: none"> • People are near-optimal and are able to make smart predictions from sparse data even with everyday cognitive judgements

→ This was an example of Bayesian updating in cognitive science.

Bayesian Models of Cognition



- Vision research has shown that sensory processing can be as accurately modeled as a process of Bayesian updating (e.g., Yuille & Kersten, 2006; Kersten, Mamassian & Yuille, 2004; or Jacobs, 2002, on Bayesian depth perception). Even motor control appears to approach Bayesian optimality (Körding & Wolpert, 2004).
- some researchers claim **that Bayesian statistics provide a general framework for understanding human inductive inference:**
- Learning (Tenenbaum, 1999), human reasoning under uncertainty (Oaksford & Chater, 1994), categorisation (Tenenbaum & Griffiths, 2001), counterfactual inference and causal representation (Pearl, 2000; Griffiths & Tenenbaum, 2006; Sloman & Lagnado, 2005), and for modeling language acquisition (e.g., Hsu & Chater, 2010).

Bayesian Models of Cognition



Bayesian Fundamentalism	Bayesian Enlightenment
<ul style="list-style-type: none"> The Bayesian model itself exists at a computational level, where its predictions are defined only based on optimal inference (Bayes laws) and decision-making. The mechanisms by which those decision are determined are outside the model's scope. 	

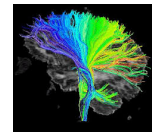
Bayesian Models of Cognition



Bayesian Fundamentalism	Bayesian Enlightenment
<ul style="list-style-type: none"> • Cognitive behaviour can be explained from rational principles alone & without reference to psychological or neurological processes <ul style="list-style-type: none"> • Significantly under constrained • Like Behaviourism (black box)? 	<ul style="list-style-type: none"> • Developed in conjunction with mechanistic considerations <ul style="list-style-type: none"> • Performing model comparisons of different Bayesian models • Take into account representations

MARR'S (1982) LEVELS

- **Computational**
 - What problem is the brain solving? What information is required? What is the structure of the environment?
- **Algorithmic**
 - What processes does the mind execute to produce the solution?
 - What algorithms are computed?
- **Implementational**
 - Hardware: How are those algorithms implemented in the brain?



TOPIC OVERVIEW

2. The Future of AI:

Human vs. Machine Learning

- INTERMEZZO: DISCUSSION

What do you think an AI has to possess in order to be truly “intelligent”?

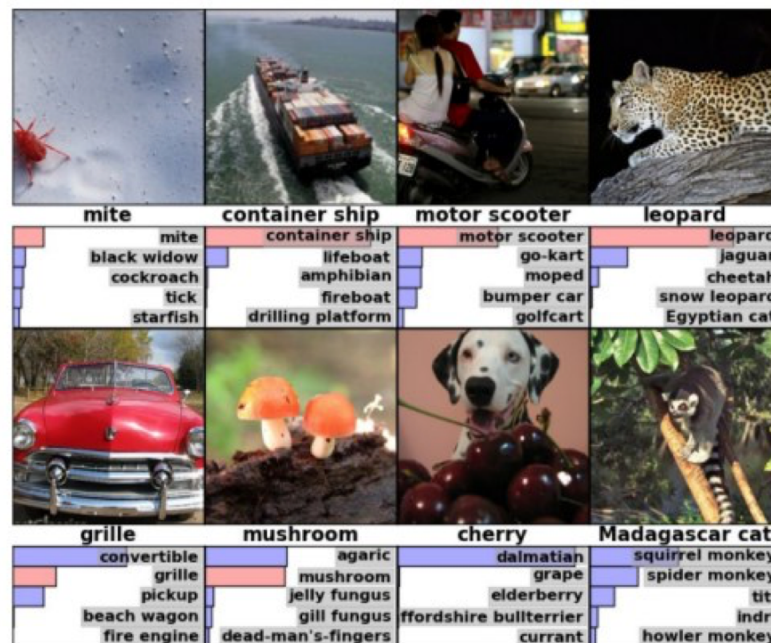
- i.e., Learning and thinking like a person

Recent developments in AI

- Engineering trends: **Deep Neural Networks**
- What can they do?
 - Object recognition
 - Speech recognition
 - Learn how to play video games
 - ...

Deep Neural Networks: object recognition

- **Krizhevsky et al. (2012)**: deep convolutional network that nearly halved the error rate of previous state-of-the-art algorithms



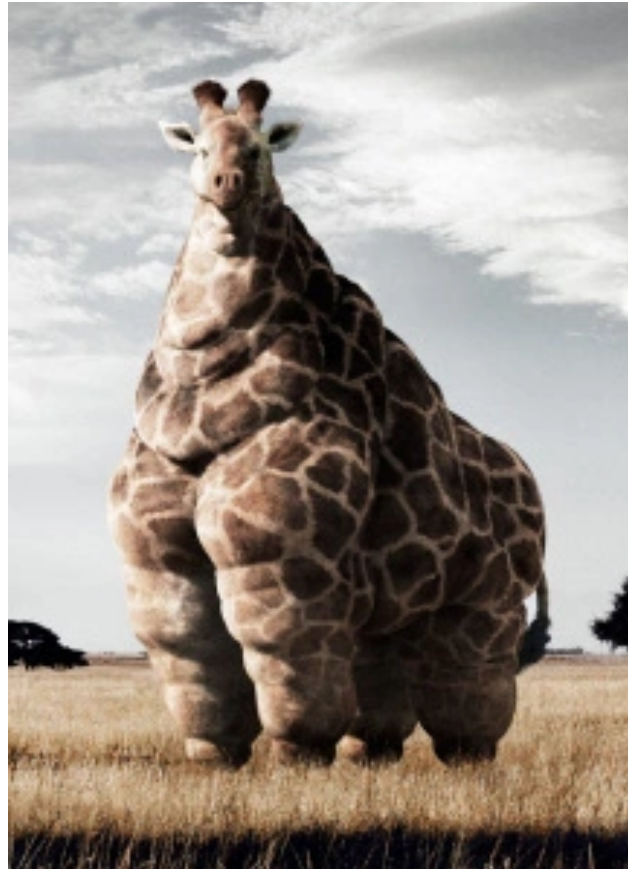
- Most challenging benchmark: ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories

How do Neural Networks (NNs) work? A very brief Intro.

Example: Object recognition

Example: Object recognition

- Try to identify:



Object recognition

- We do this with the 80 billion neurons in our brain working together to transmit information.
- This remarkable system of neurons is also the inspiration behind a widely-used machine learning technique called *Artificial Neural Networks (NN)*.



Object recognition

- Some computers using NN's have even outperformed humans in recognizing images.



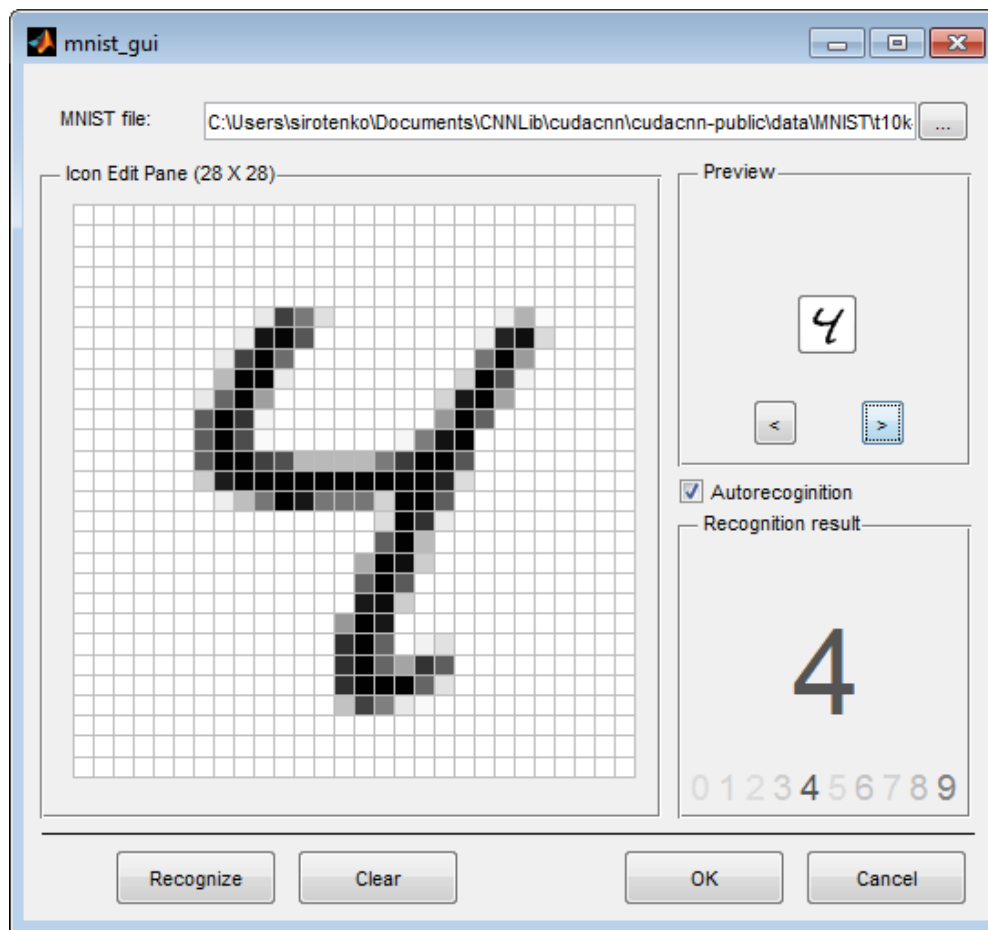
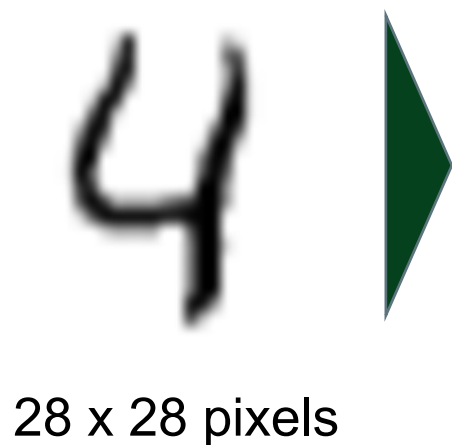
The Problem: Recognize handwritten digits

- Image recognition is used in technology for visual surveillance, guiding autonomous vehicles, identifying abnormalities in X-ray images, smartphone app that converts handwriting into typed words, etc.



- How can we train an artificial neural network to recognize images of handwritten digits?

The Problem: Classify digits 0-9



Input layer:
784 variables,
one per pixel

784 pixels coded as numbers
based on darkness

The Problem: Classify digits 0-9

- 1. Training:** NN model is trained by giving it examples of 10,000 handwritten digits, together with the correct digits they represent. → allows the NN model to understand how the handwriting translates into actual digits.



Handwritten digits in the MNIST database

- 2. Testing/Validation:** After the NN model is trained, we can test how well the model performs by giving it 1,000 new handwritten digits without the correct answer. The model is then required to recognize the actual digit.
- 3. This process is called *cross-validation*.**





The Problem: Recognize handwritten digits

Contingency table to view results:

		Predicted Digit											
		0	1	2	3	4	5	6	7	8	9	Total	%
Actual Digit	0	84	0	0	0	0	0	1	0	0	0	85	99
	1	0	125	0	0	0	0	1	0	0	0	126	99
	2	1	0	105	0	0	0	0	4	5	1	116	91
	3	0	0	3	96	0	6	0	1	0	1	107	90
	4	0	0	2	0	99	0	2	0	2	5	110	90
	5	2	0	0	5	0	77	1	0	1	1	87	89
	6	3	0	1	0	1	2	80	0	0	0	87	92
	7	0	3	3	0	1	0	0	90	0	2	99	91
	8	1	0	1	3	1	0	0	2	81	0	89	91
	9	0	0	0	0	1	0	0	6	2	85	94	90
Total		91	128	115	104	103	85	85	103	91	95	1000	-

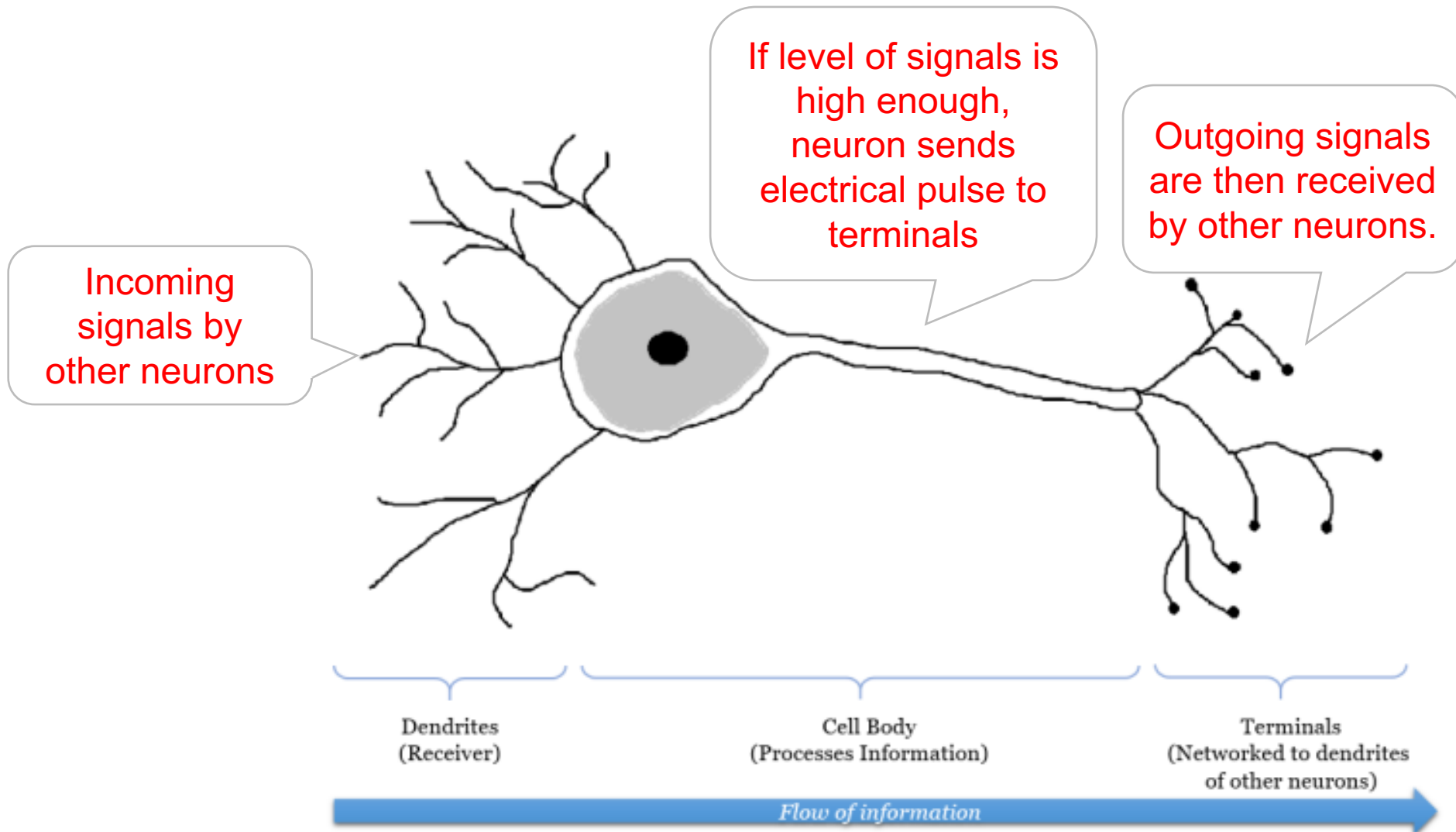
- Out of the 1,000 handwritten images that the model was asked to recognize, it correctly identified 922 of them, which is a 92.2% accuracy.

The Problem: Recognize handwritten digits

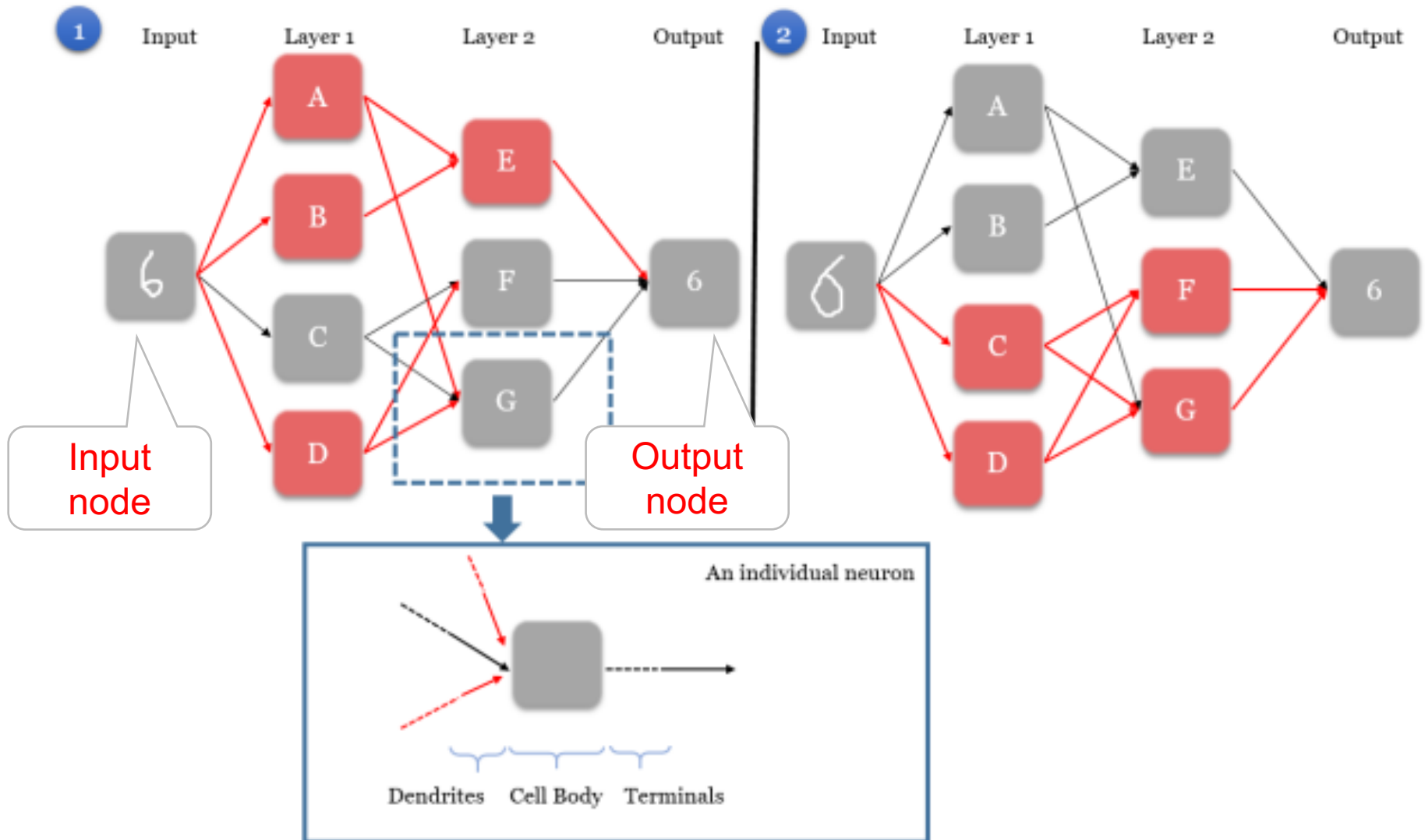
	<p><u>Prediction:</u> Digit 7 – 99% Digit 3 – 1%</p>		<p><u>Prediction:</u> Digit 7 – 94% Digit 2 – 5% Digit 3 – 1%</p>
	<p><u>Prediction:</u> Digit 8 – 48% Digit 2 – 47% Digit 3 – 4% Digit 1 – 1%</p>		<p><u>Prediction:</u> Digit 8 – 58% Digit 2 – 27% Digit 6 – 12% Digit 0 – 2% Digit 5 – 1%</p>

- Some of the digits get confused.

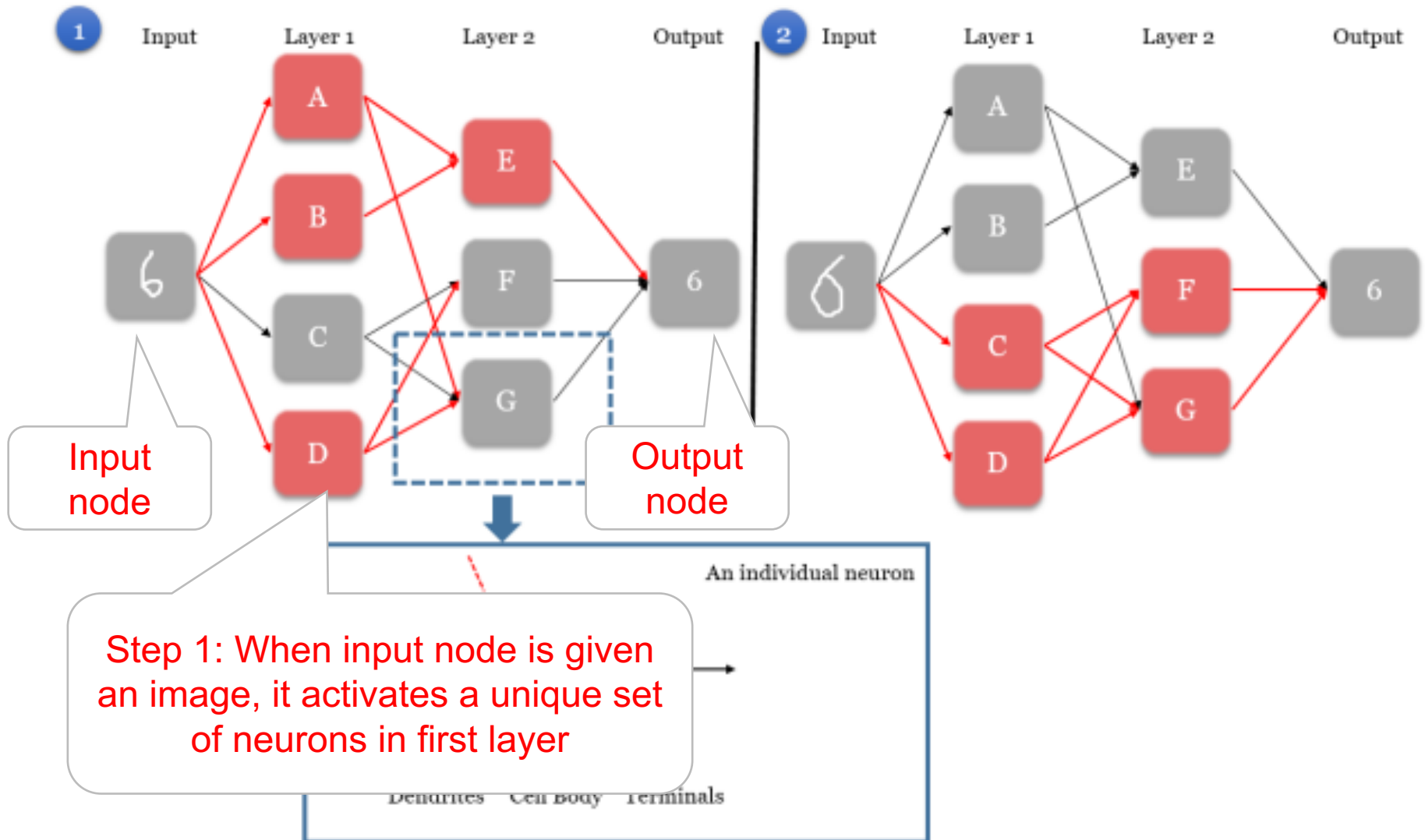
The neurons that inspired the network



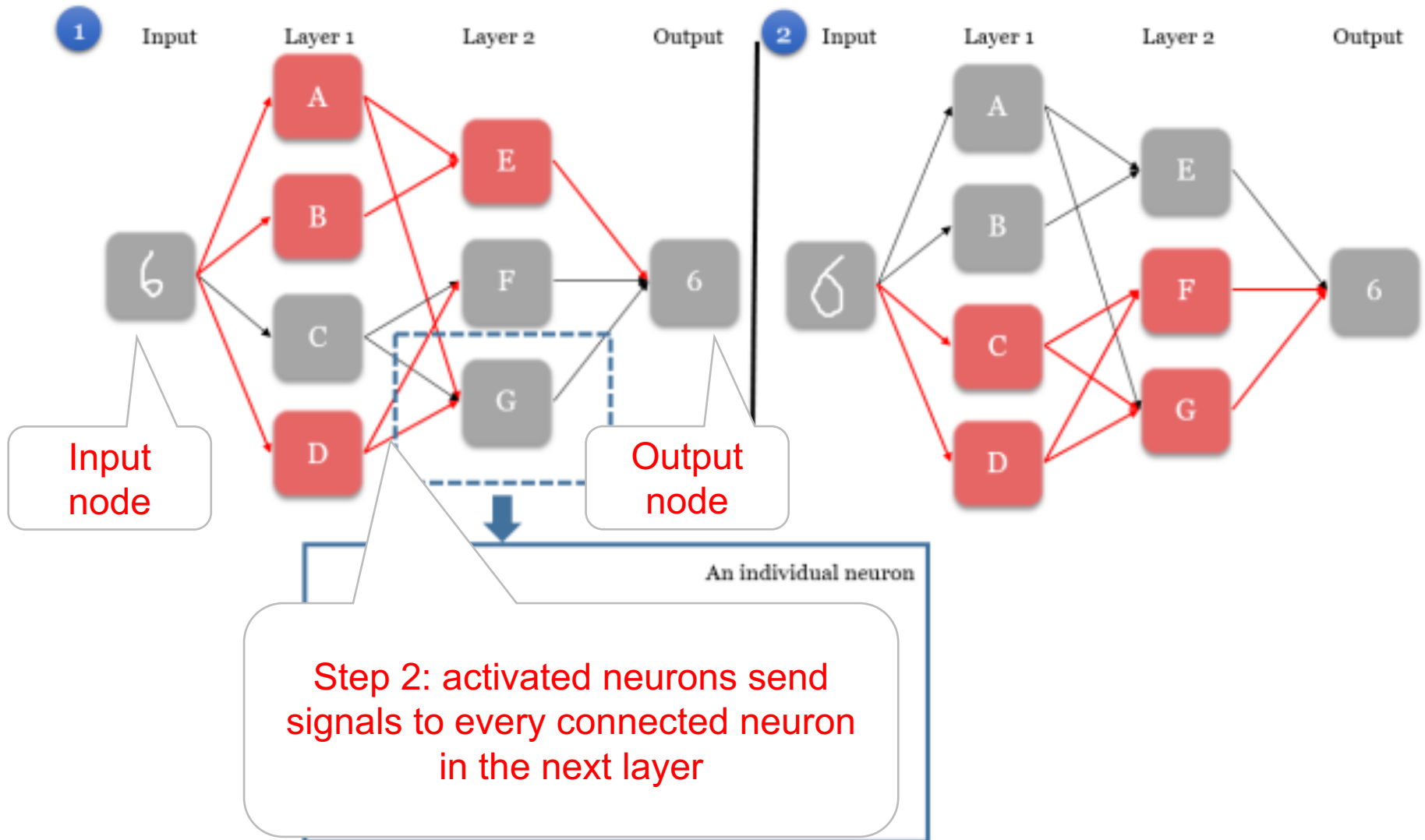
HOW THE MODEL WORKS



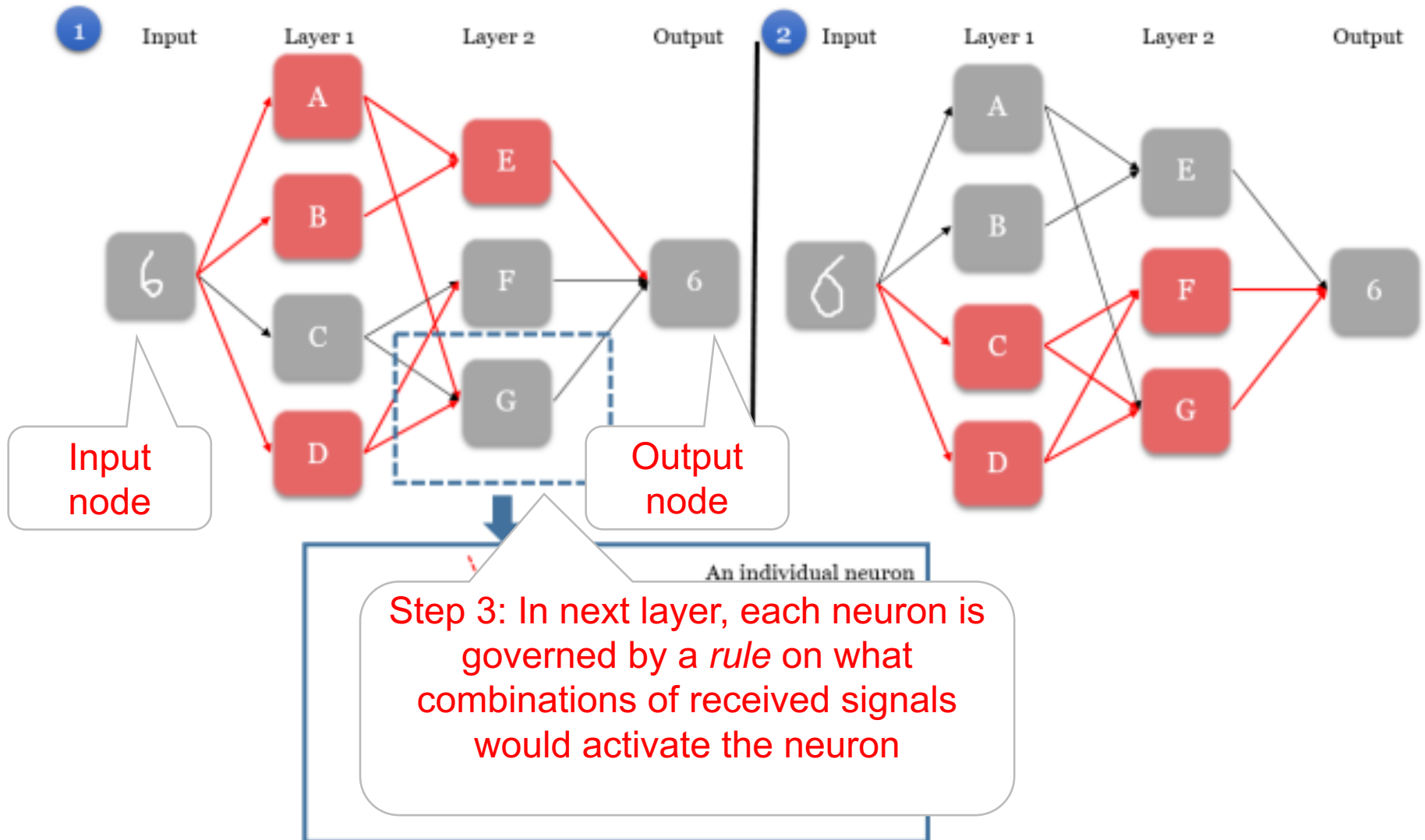
HOW THE MODEL WORKS



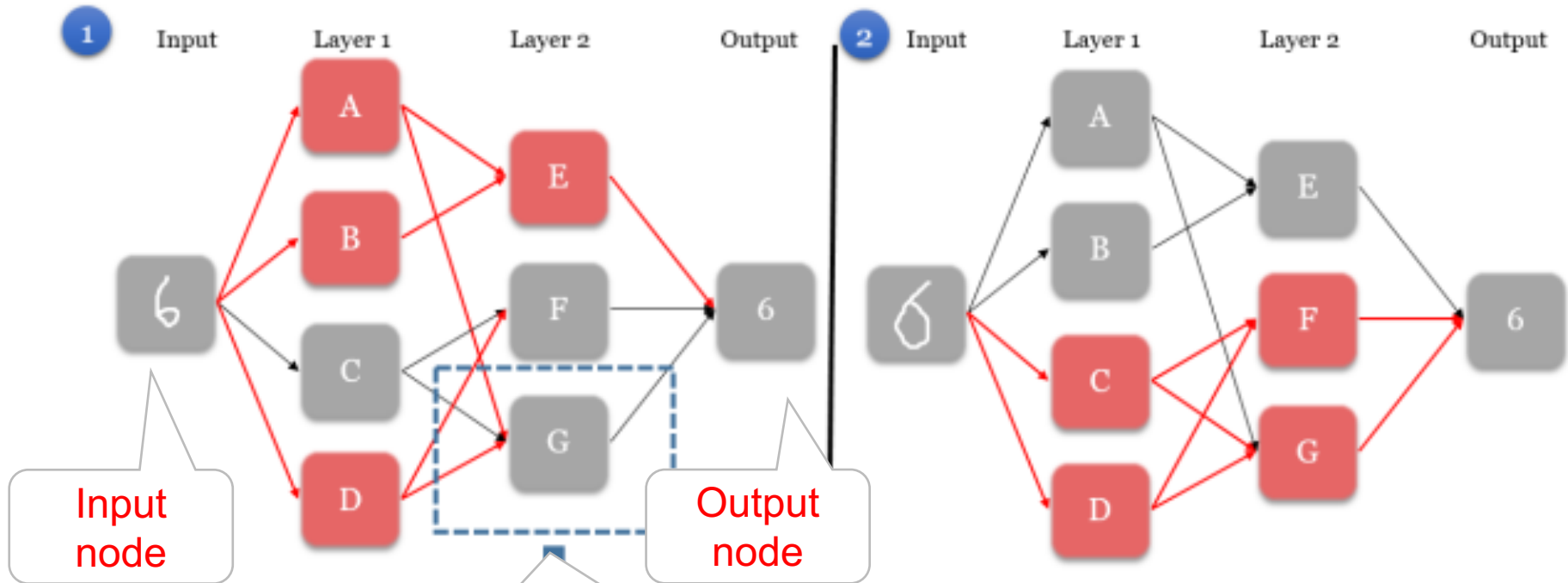
HOW THE MODEL WORKS



HOW THE MODEL WORKS

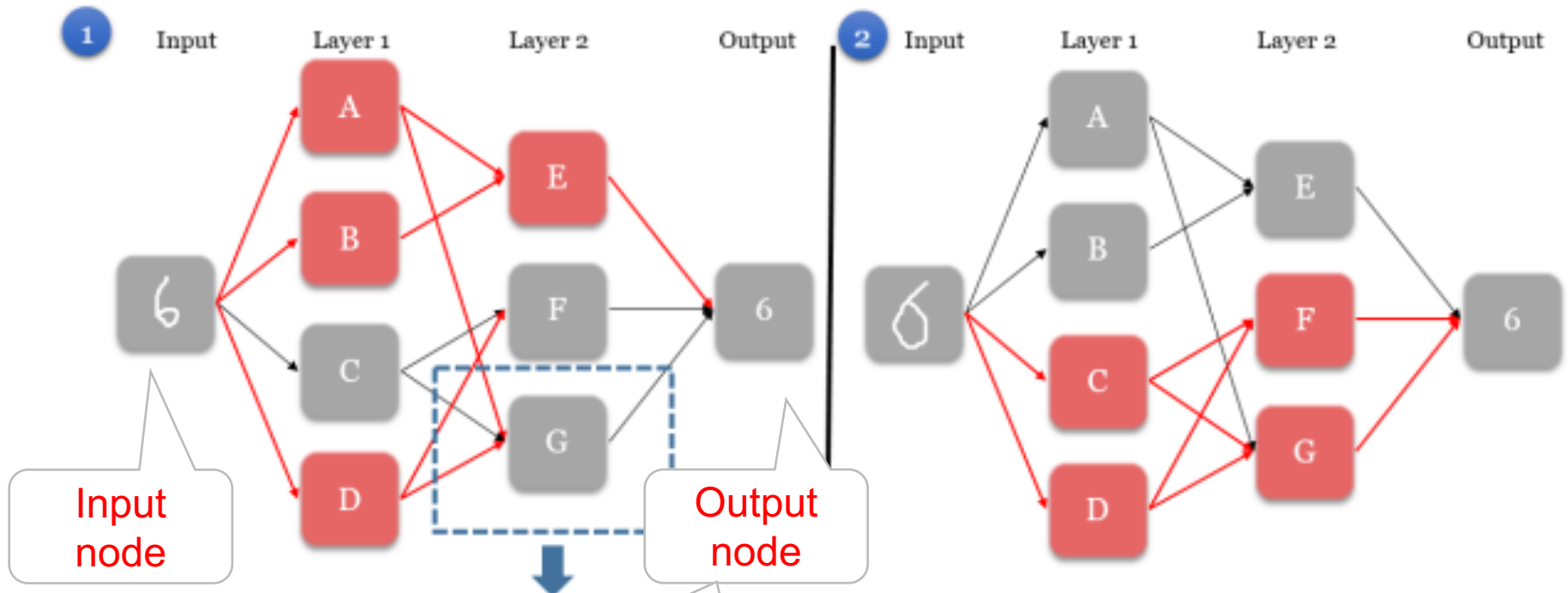


HOW THE MODEL WORKS



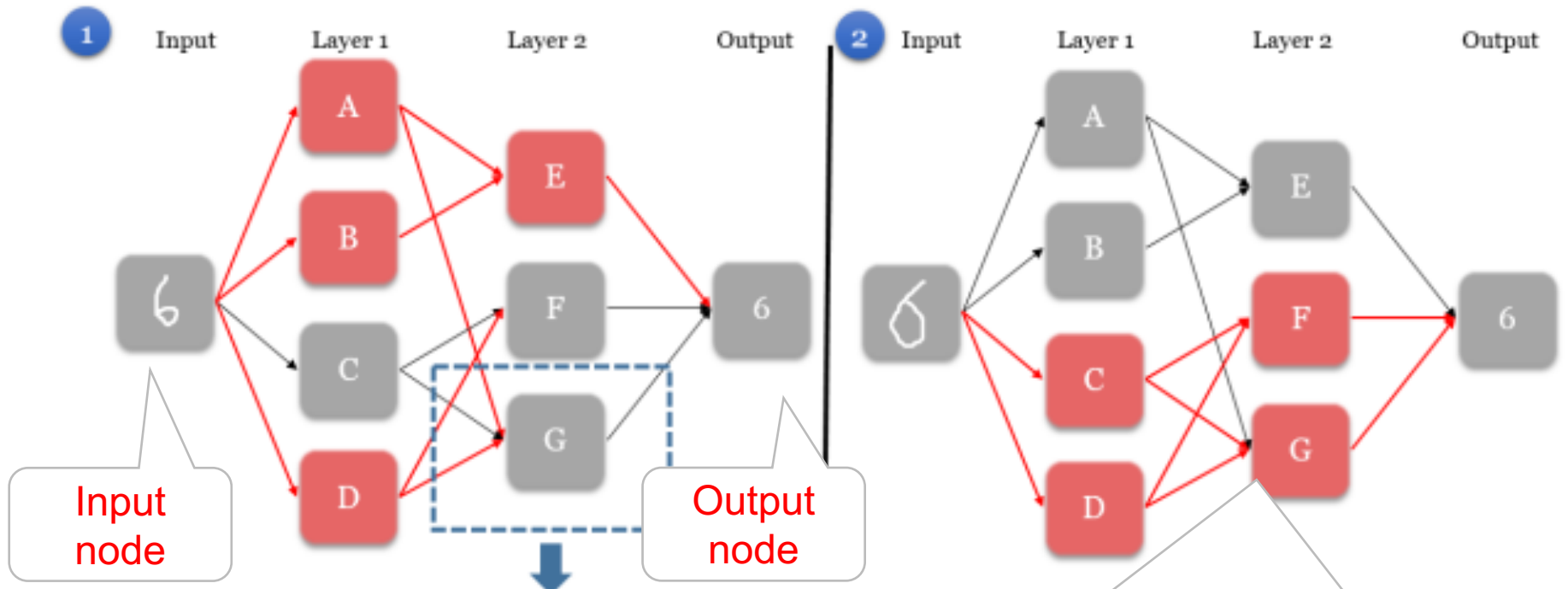
Step 4: Steps 2-3 are repeated for all the remaining layers (it is possible for the model to have more than 2 layers), until we are left with the output node.

HOW THE MODEL WORKS



Step 5: output node deduces the correct digit based on signals received from neurons in the layer directly preceding it (layer 2).

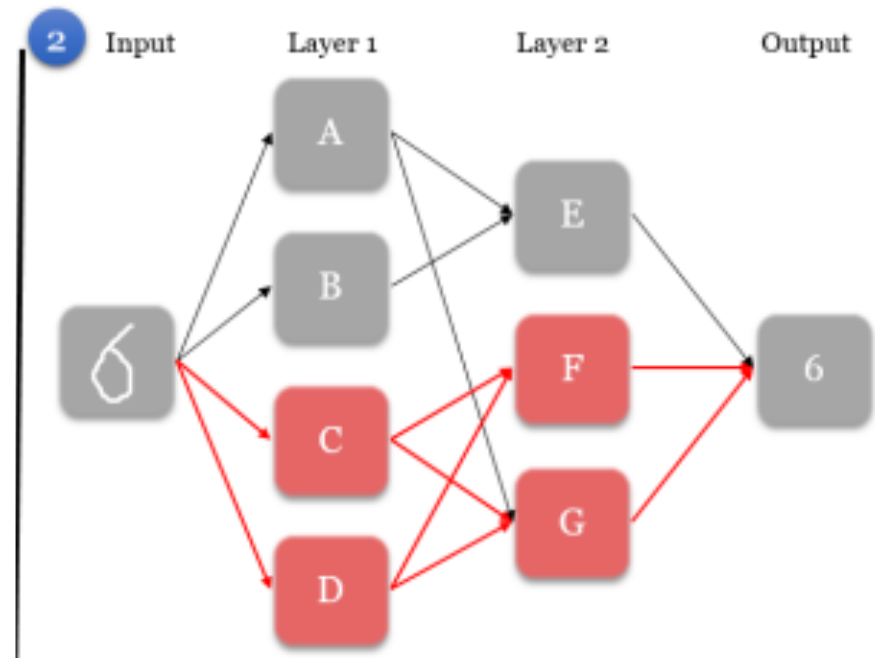
HOW THE MODEL WORKS



Each solution in Scenario 1 and 2 can be represented by different combinations of activated neurons. Because the images fed as input in Scenarios 1 & 2 are different, the network activates different neural paths from input to the output. However, the output still recognizes both images as the digit “6”.

Training the model

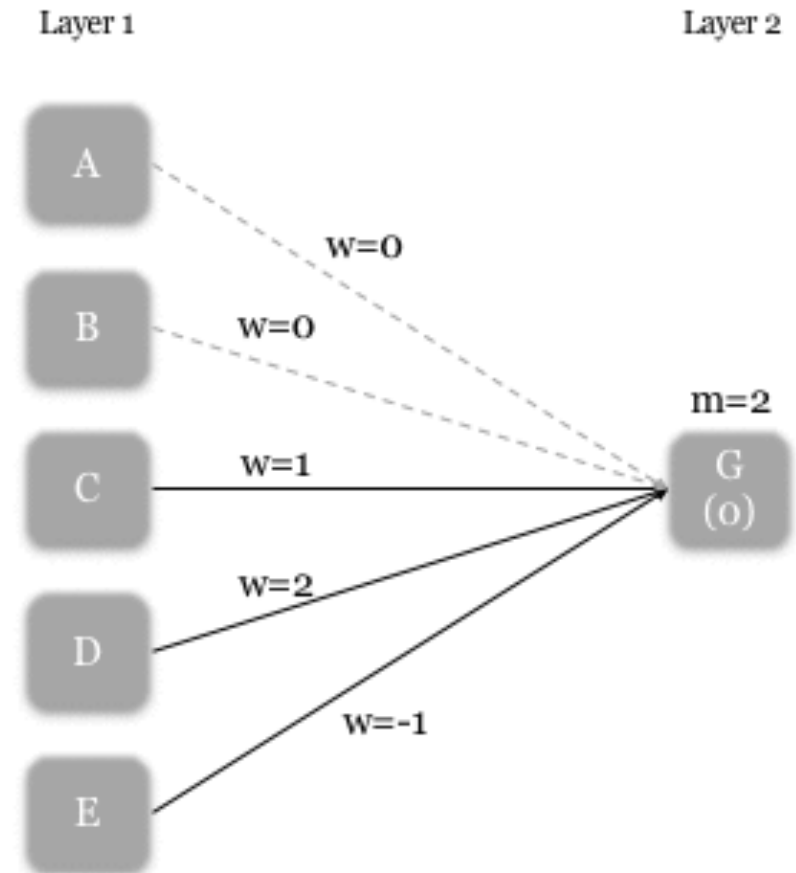
- Decide on number of layers and number of neurons in each layer
- For the digit recognizer NN, 3 layers with 500 neurons each were used.
- Key factors involved in training a model are:
 - metric to evaluate the model's accuracy, i.e., loss function (SSE)
 - Rules that govern whether neurons are activated or not



Training the model

Neuron's activation rule has 2 components:

1. the weight (i.e. strength) of incoming signals [w]
2. minimum received signal strength required for activation [m].



Rules for neuron G

Neural Networks: Limitations

1. Computationally expensive.

Training a NN takes more time and CPU power compared to training other types of models (e.g., random forests)

Although NN are not a new technique (1980s and 1990s), they have had a revival in recent years because of hardware advances

2. Lack of feature recognition.

The NN is unable to recognize images if they take on slight variations in shape, or are placed in a different location.



Advanced NN: Convolutional Neural Networks

Neural Networks



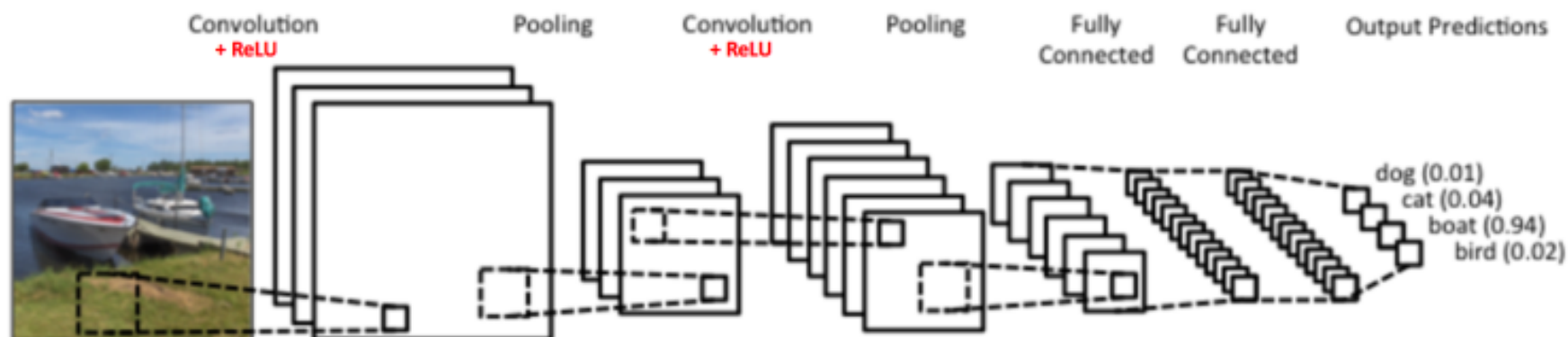
Convolutional Neural Network

Solves the problem of **lack of feature recognition**, by **looking at various regions of the image**.



Convolutional Neural Networks (CNNs)

- **CNNs are more efficient and widely used** = A neural network that uses a trainable filter instead of fully-connected layers with independent weights.



Convolutional Neural Networks (CNNs)

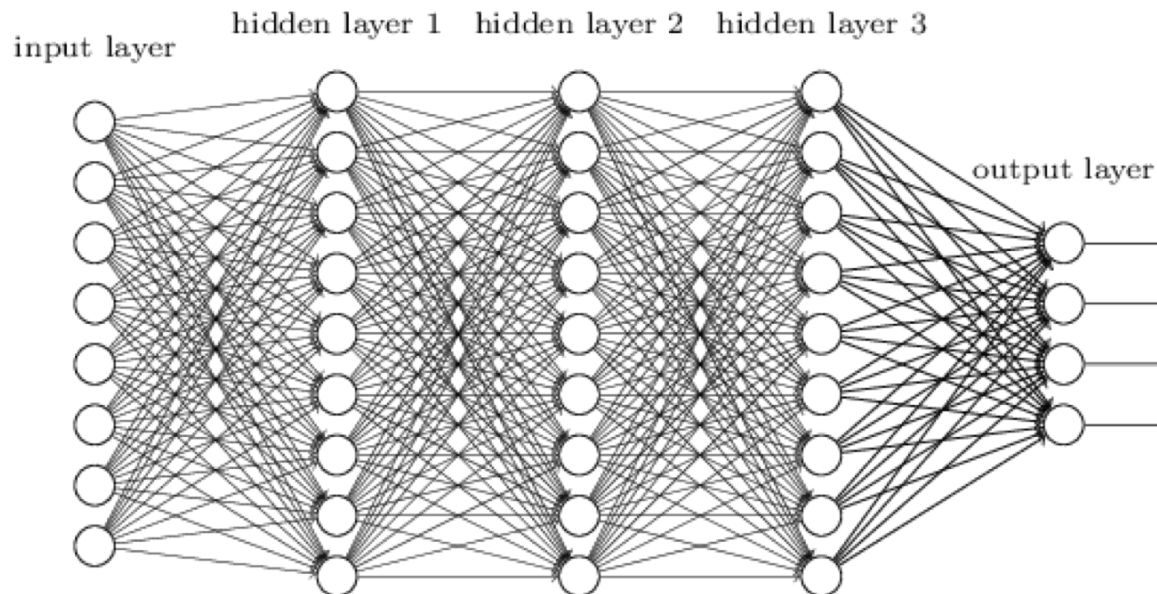
- **CNNs are more efficient and widely used** = A neural network that uses a trainable filter instead of fully-connected layers with independent weights.



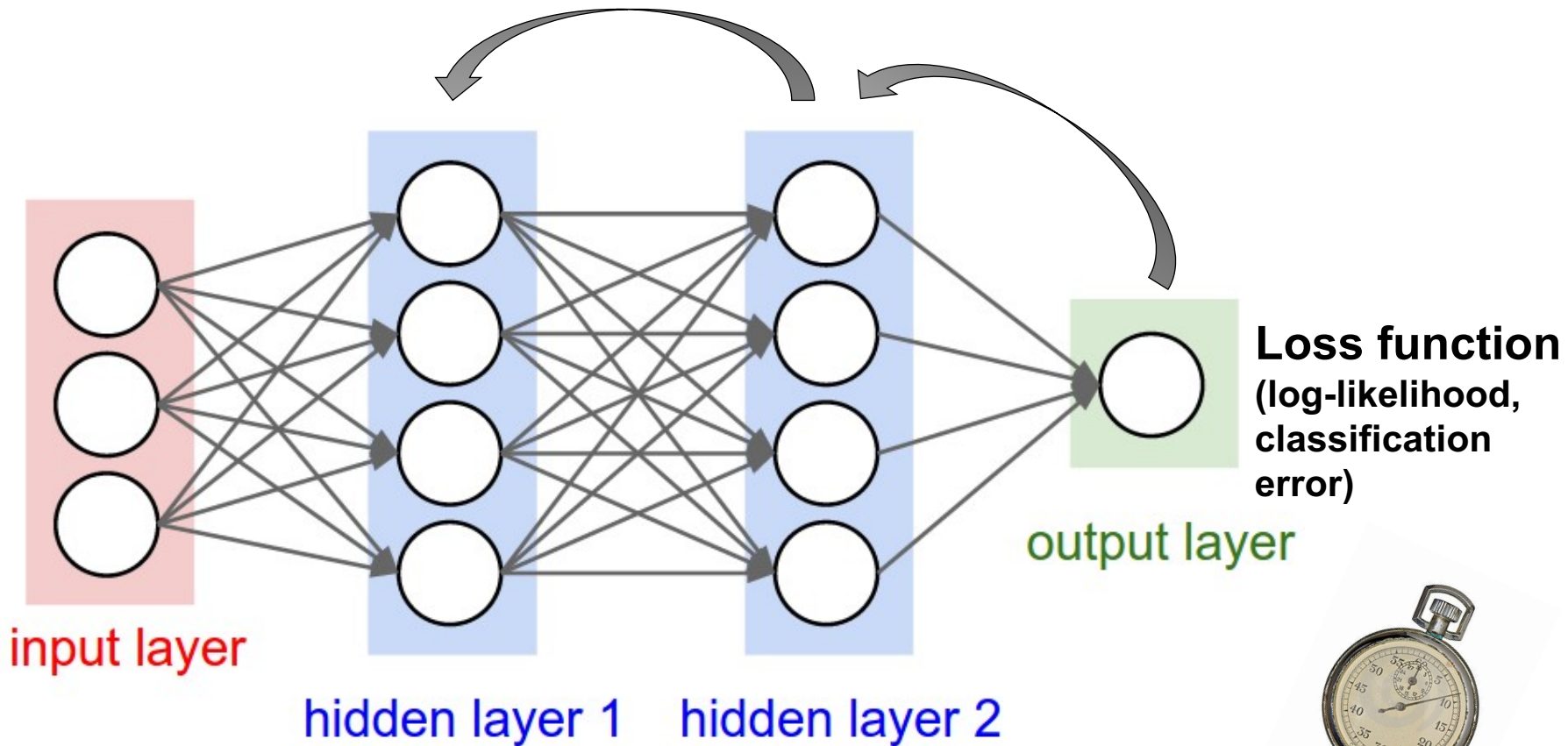
Input

Advanced NNs: Deep Neural Networks

Deep Neural Networks = A neural network with at least one hidden layer (some networks have dozens).

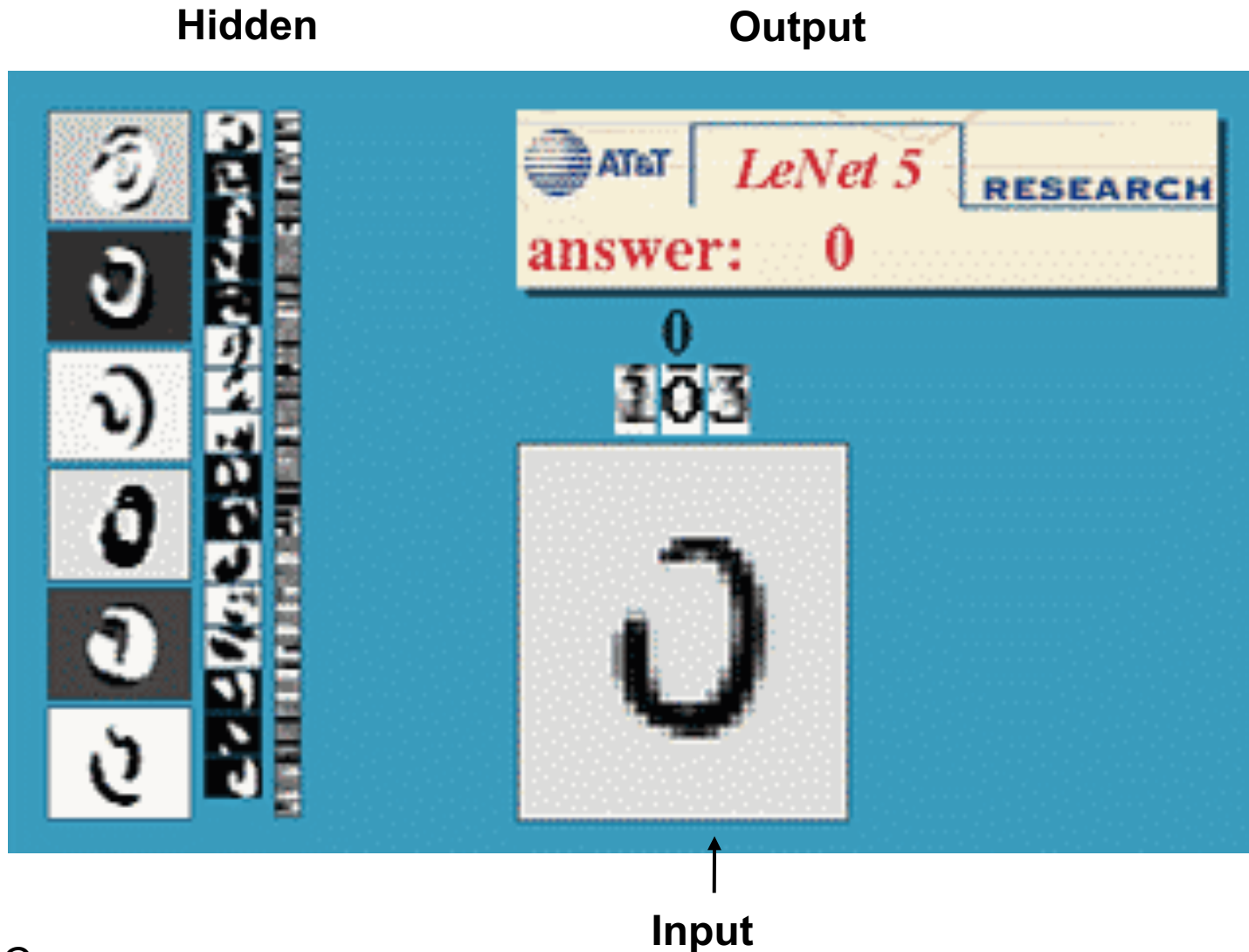


Optimization Method ~ Backpropagation



Run until error stops improving = converge

Demo

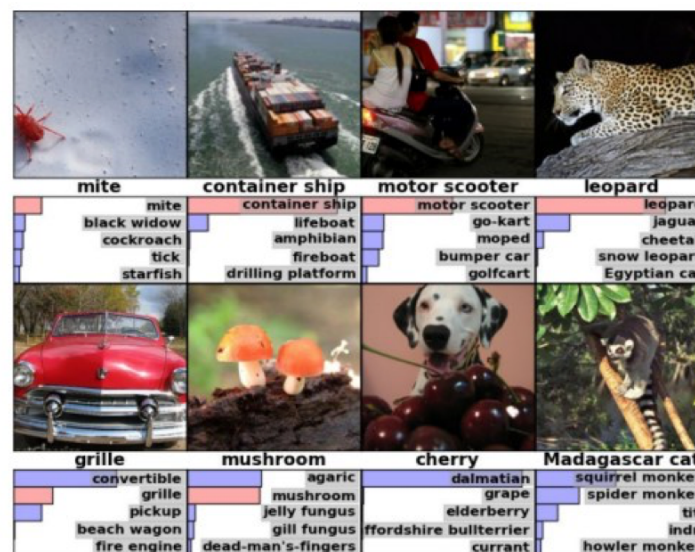


*Yann LeCun

Putting it all back together: Deep Convolutional Networks

Krizhevsky et al. (2012): a deep convolutional neural network is capable of achieving record breaking results in object recognition using purely supervised learning.

“It is notable that our network’s performance degrades if a single convolutional layer is removed. “



Deep Convnets continue to dominate

- Best results achieved are very close to human-level performance at an error rate of 0.2% (Ciresan, Meier & Schmidhuber, 2012).
- **He, Shang, Ren & Sun (2015)**: recently approaching human-level performance on some benchmarks

Agriculture with deep learning drone algorithms



Dramatic improvements in:

- **Crop and tree counting**
- **Early disease detection**
- **Livestock tracking**
- **Intelligent pesticide delivery**
- **Yield forecasting**
- **Crop design**

Urban insights with Deep Learning Drone Tech



Fully automated

- Parking spot locator
- Car counting
- Optimizing traffic
- Crowd analysis

Automatic speech recognition



- Hidden Markov Models have been replaced piece by piece with deep learning

Another grand challenge of AI

- Deep Neural Networks were used by Google in Oct 2015 and Mar 2016 to defeat human champions in the game of Go



A typical argument has been that neural networks have achieved this success by virtue of their brain-like computation and ability to emulate human learning and human cognition.

However, ...

Artificial intelligence vs. Human Intelligence

- Despite performance achievements, these AI systems differ from human intelligence in many ways e.g.,
 - What they learn
 - How they learn

Discussion Intermezzo

WHAT ARE THEY MISSING?

Endowing AI with cognitive science & psychology

Intuitive physics

Theory of mind

Discovery of new
mental models =
Learning-to-learn

**Neural
Networks +**

Causal
reasoning

Inductive biases

More structure

Endowing neural networks with cognitive science

- **Argument (Lake et al., *in press*):** As long as natural intelligence remains the best example of intelligence, we believe that the project of **reverse-engineering** the human solutions to difficult computational problems will continue to inform and advance AI.

! Important Distinction: Of course not all AI is built for the purpose of taking neural inspiration and make claims of cognitive and neural plausibility. (e.g., the Automated Statistician)

Building human-like learning and thinking machines

Lake, Ullman, Tenenbaum & Gershman (2017):

1. Developmental ingredient: **intuitive physics**
2. Developmental ingredient: **intuitive psychology**
3. Learning ingredient: *model building* (**causal models**)
4. Learning ingredient: **compositionality**
5. Learning ingredient: **learning-to-learn**
6. Learning ingredient: **speed**
(our minds are able to build rich models instantly)

Core
ingredients
of human
intelligence

Why focus on development?

- “If an ingredient is present early in development, it is certainly active and available well before a child or adult would attempt to learn the types of tasks discussed here. Thus is true regardless of whether the early-present ingredient itself is learned from experience or innately present.” (Lake et al., *in press*)
- The earlier an ingredient is present, the more likely it is to be foundational to later development and learning.

Building human-like learning and thinking machines

An Important Distinction:

Generic neural networks	Human learning mechanisms
<ul style="list-style-type: none"> • aimed at pattern recognition • Not aimed at model building • Don't know how to draw inferences • "model-free" 	<ul style="list-style-type: none"> • aimed at rich model-building <ol style="list-style-type: none"> 1. Learn causal model of the task 2. Then use model to plan action sequences • Know how to draw inferences • "model-based"



Building human-like learning and thinking machines

Generic neural networks	Human learning mechanisms
<ul style="list-style-type: none"> • aimed at pattern recognition • Not aimed at model building • Don't know how to draw inferences • "model-free" 	<ul style="list-style-type: none"> • aimed at rich model-building <ol style="list-style-type: none"> 1. Learn causal model of the task 2. Then use model to plan action sequences • Know how to draw inferences • "model-based"



Can we teach neural networks to draw inferences?

Challenges for building more human-like machines

- 1) The Characters Challenge
 - Learning simple visual concepts (simple for humans!)

1) The Characters Challenge

- Similar to **MNIST benchmark** for recognizing handwritten digits

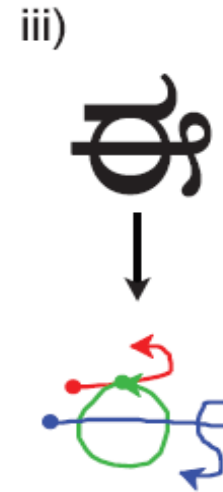
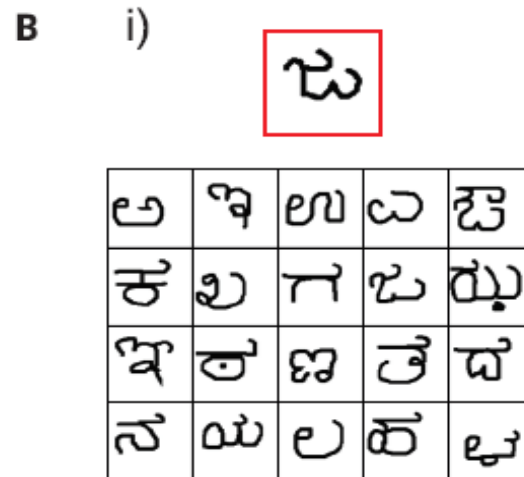
A handwritten digit '4' in black ink, slightly tilted and with a casual, human-like appearance.

→ Just because humans and neural networks perform equally well on the MNIST recognition task, does not mean they learn and think the same way:

- 1) People learn from fewer examples
- 2) People learn richer representations

1) Learning handwritten characters

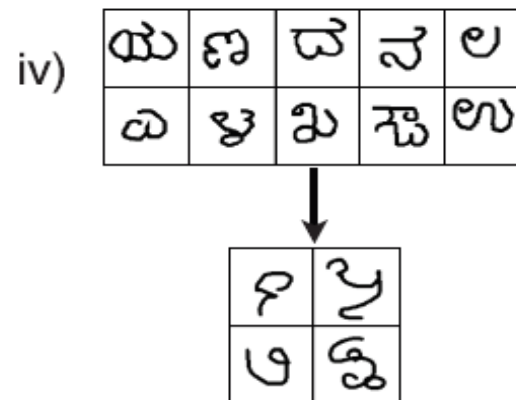
1. People can learn to recognize a new handwritten character from a single example



3. People can parse a character into its most important parts and relations.

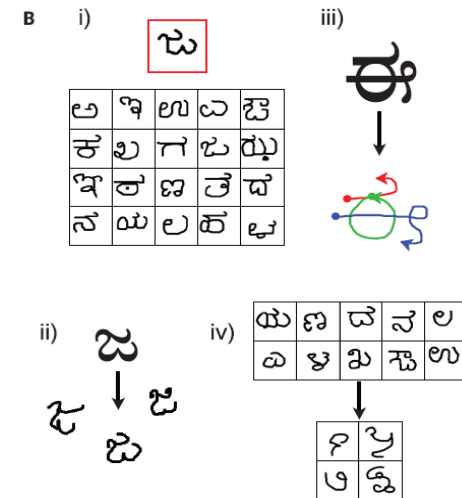
People learn more than how to do pattern recognition: they learn a concept.

2. In addition to recognizing new examples, people can also generate new examples.



4. People can generate new characters given a small set of related characters.

Interim conclusions



- ❖ People's additional abilities come for free along with the acquisition of the **underlying concept**.
- ❖ People can build **models** and then use them for arbitrary new tasks and goals.
- ❖ People **learn a lot more from a lot less** – and capturing these human-level learning abilities in machines is the challenge.

Core Ingredient 1: Intuitive physics



PHSCS 121

Principles of Physics 1

- Young children have rich knowledge of intuitive physics
- At 2 months and earlier, infants expect inanimate objects to follow principles of persistence, continuity, cohesion and solidity.
- At 6 months, infants have developed different expectations for solid bodies, soft bodies, and liquids.
- By their 1st birthday, infants have gone through several transitions of comprehending basic physical concepts such as inertia, containment and collisions

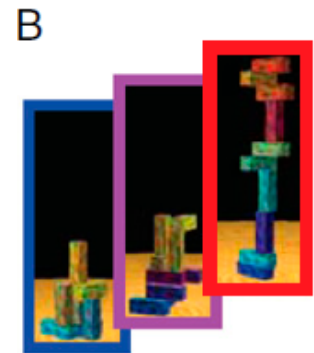
**All influences
later learning..**

Core Ingredient 1: Intuitive physics

- There is no single agreed-upon computational account of these early physical principles and concepts
 - Decision trees? (Baillargeon et al., 2009)
 - Lists of rules? (Sigler & Chen, 1998)
 - **Intuitive physical reasoning as inference over a physics software engine**, i.e., the kind of simulators that power modern-day animations and games (Bates et al., 2015, Battaglia et al., 2013; Gerstenberg et al., 2015)

Intuitive physical reasoning as inference over a physics software engine

- Hypothesis: People reconstruct a perceptual scene using internal representations of the objects and their physically relevant properties (mass, elasticity, surface friction), and forces acting on the objects (gravity, friction, collision impulses).
- intuitive physical state representation is approximate and probabilistic, and oversimplified. → But it can support mental simulations that can predict how objects will move in the immediate future in response to forces or on their own.



Core Ingredient 1: Intuitive physics

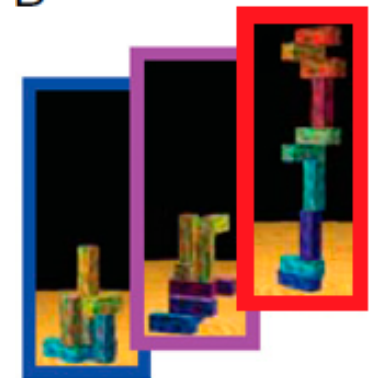
- http://www.ucl.ac.uk/lagnado-lab/neil_bramley.html

Integration 1: Intuitive physics + Neural Networks

- 1) making predictions without simulating physics: Facebook AI researchers (Lerer et al., 2016) trained a deep convnet to predict the stability of block towers (“Jenga”) from simulated images (physical judgment by PhysNet)
- 2) Alternatively, could a neural network be trained to emulate a general-purpose physics simulator, given the right type and quantity of training data, such as the raw input experienced by a child?

Challenges: what will the deeper layers encode?

How will the deep net generalize?



Intuitive psychology



- Pre-verbal infants distinguish animate agents from inanimate objects
- Infants also expect agents to act contingently and reciprocally, to have goals, and to take efficient actions towards those goals subject to constraints (at 3 months, **anti-social** and **anti-cooperative**)
- One possibility is that intuitive psychology is “**hard-coded** and “**shut down**” (Schlottmann et al., 2013)
- Or: generative models of action choice = Formalize concepts such as ‘goal’, ‘agent’, ‘planning’, ‘cost’, used to describe core psychological reasoning in infancy

Applying deep networks here is also new.
→ It could learn visual cues, heuristics and summary stats of a scene with agents.

Core Ingredient 3: Model building

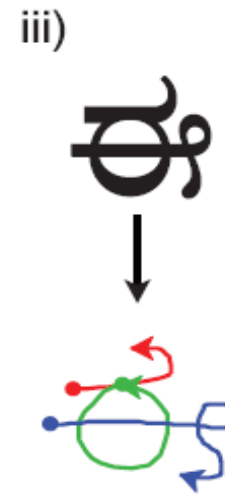
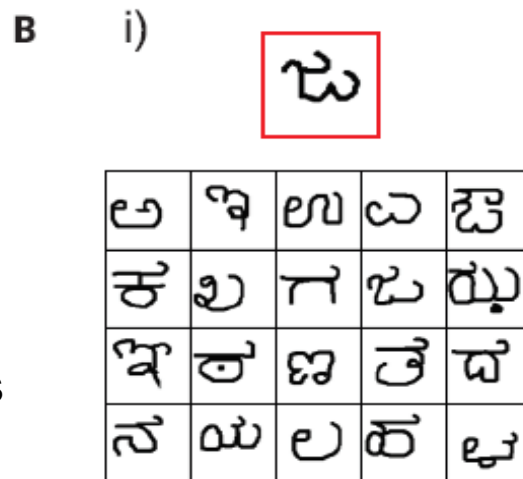
- Deep neural networks are extremely data hungry.
- **Rapid model building:** When children learn in their native language, they make meaningful generalizations from only a few examples of a new concepts such as *hairbrush* or *pineapple*
- Why are NNs so much less efficient? (not always of course)
- Even with just a few examples, people can learn remarkably rich conceptual models.

1) Learning handwritten characters

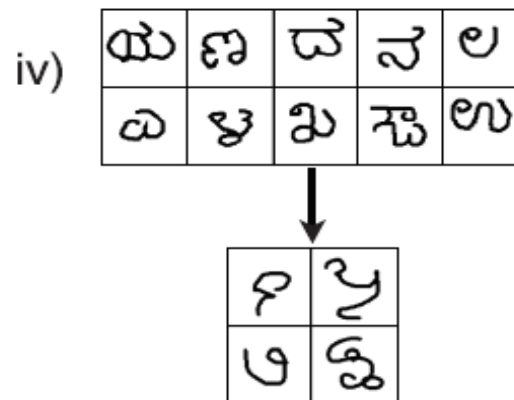
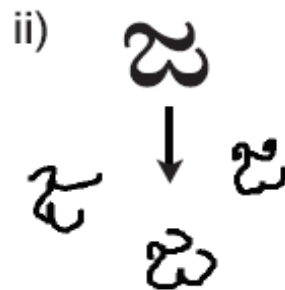
1. People can learn to recognize a new handwritten character even if someone else draws it!

People learn more than how to do pattern recognition: they learn a concept.

2. In addition to recognizing new examples, people can also generate new examples.



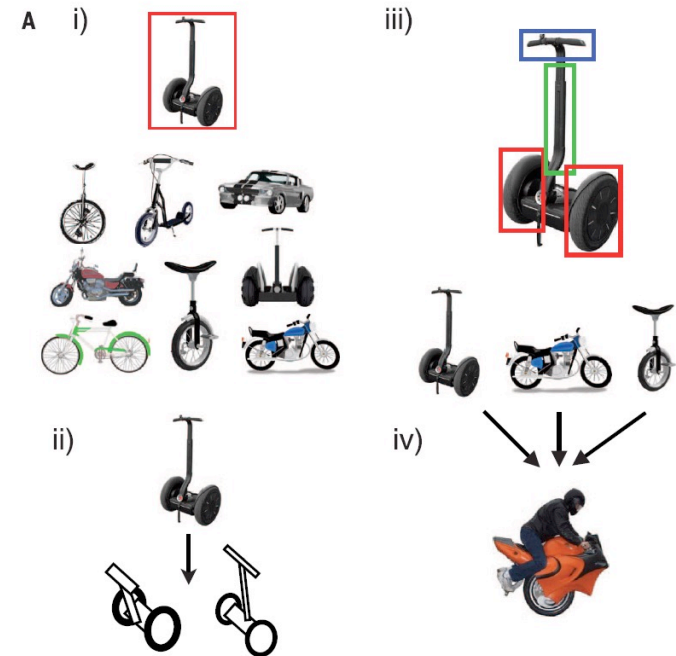
3. People can parse a character into its most important parts and relations.



4. People can generate new characters given a small set of related characters.

Core Ingredient 3: Models allow

- Concepts support classification
- Concepts support prediction
- Action
- Communication
- Imagination
- Explanation
- Composition



- => These abilities are not independent of the model, rather they hang together and come for free with the acquisition of the underlying concept.

Can we imitate this rapid model building?

- ❖ Conclusions on the Characters Challenge by Lake et al. (2015)'s probabilistic program induction:
- ❖ While both people and model represent characters as a sequence of pen strokes and relations, **people have a far richer repertoire of structural relations between strokes.**
- ❖ People can efficiently integrate across multiple examples of a character combining different variants into single coherent representation.
- ❖ **-> AI still not as good as human performance!**

Can we imitate this rapid model building?

- ❖ Lake et al. (2015)'s probabilistic program induction:

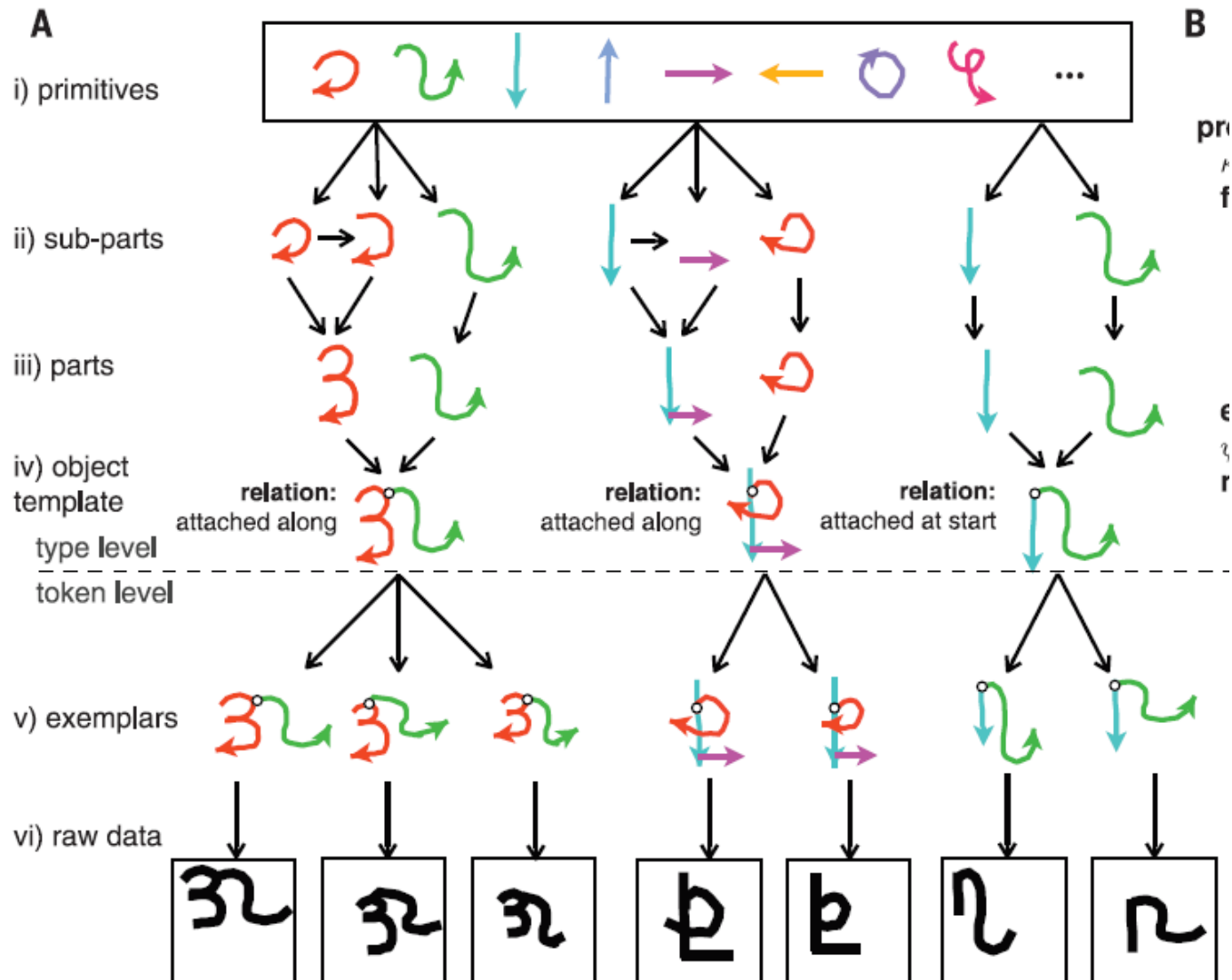


People




















Machines




















- Compare the ability to learn new handwritten characters from the world's alphabets.
 - Deep learning
 - **Bayesian Program learning (BPL)** represents concepts as simple stochastic programs, i.e., structured procedures that generate new examples of a concept when executed




















A causal, compositional model of handwritten characters






















Probabilistic inference allows the model to generate new examples from just one example of a new concept

	1			<div style="border: 1px solid black; padding: 2px; display: inline-block;">  </div>	2		
							
							
							

	1			<div style="border: 1px solid black; padding: 2px; display: inline-block;">  </div>	2		
							
							
							

	1			<div style="border: 1px solid black; padding: 2px; display: inline-block;">  </div>	2		
							
							
							

	1			<div style="border: 1px solid black; padding: 2px; display: inline-block;">  </div>	2		
							
							
							

Visual Turing Test: Human or Machine?

M

	1				2		
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M

M

	1				2		
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M

M

	1				2		
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M

M

	1				2		
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M
M	M	M	M	M	M	M	M



The Competition:

Lake, Salakhutdinov, Tenenbaum (2015)

- ❖ BPL's learning structure can do one-shot classification at human-level performance & outperform current convolutional deep learning models
- ❖ Representations that BPL learns also enable it to generalize in other, more creative human-like ways (visual Turing Test)
- ❖ While both people and model represent characters as a sequence of pen strokes and relations, **people have a far richer repertoire of structural relations between strokes.**

i)

ಓ

ಓ	ಇ	ಉ	ಊ	ಋ
ಈ	ಁ	ಂ	ಃ	಄
ಏ	ಐ	ಊ	ಋ	ಠ
ಡ	ಢ	ಣ	ಠ	ಡ

The 3 main ingredients behind the success of this model:

- 1) Compositionality
- 2) Causality
- 3) Learning-to-learn

The 3 main ingredients behind the success of this model:

1) Compositionality

= New representations can be constructed through the combination of primitive elements.

** Eric's lecture will dive into this more.



The 3 main ingredients behind the success of this model:

1) Causality

= Explaining observed data through the construction of *causal* models of the world

- Generative process of BPL model for characters resembles the causal steps of writing in the world

** Dave's + Christos' lecture will dive into this more.

The 3 main ingredients behind the success of this model:

1) Learning-to-learn

= Theory-based inference

** Active Learning lecture will dive into this more.



The 3 main ingredients behind the success of this model:

- 1) Compositionality
- 2) Causality
- 3) Learning-to-learn

- Explaining observed data through the construction of *causal* models of the world
- Hallmark of human-level learning (Lake et al., in press)

Discussion Intermezzo

WHAT IS MISSING?

Conclusions Part II

- 1. Richness and flexibility of human learning suggests that learning as model building is better metaphor than learning as pattern recognition**
- 2. Human capacity for one-shot learning suggests that these models are built upon rich domain knowledge rather than starting from blank slate.**
- 3. In contrast, much of recent progress in deep learning has been on pattern recognition problems, incl. object recognition, speech recognition, and (model-free) video game learning, which use large data sets but little domain knowledge.**

References

- Griffiths & Tenenbaum (2006)
- Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32(01), 69-84.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(04), 169-188.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. arXiv preprint arXiv:1604.00289. <https://arxiv.org/abs/1604.00289>