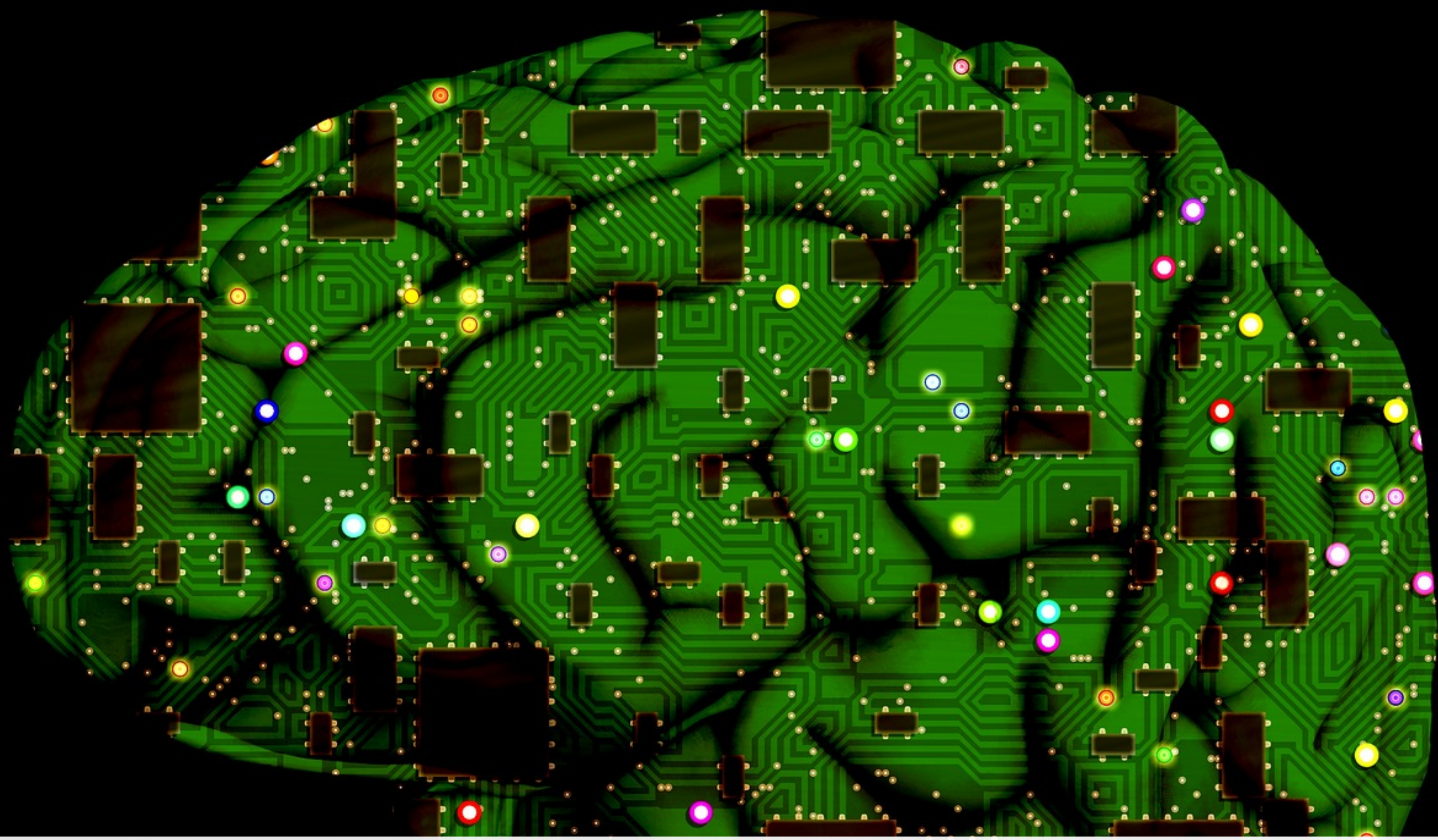
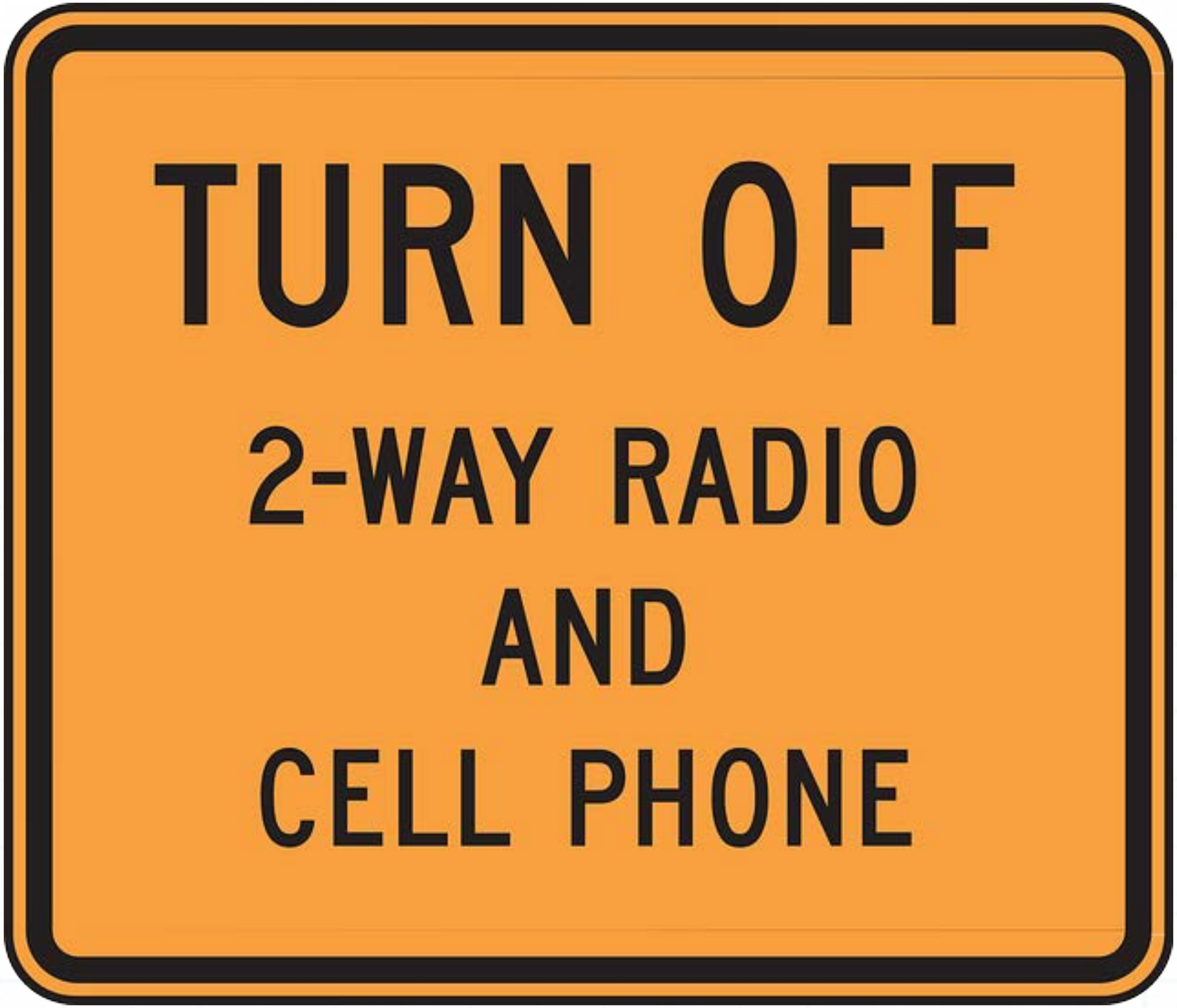


An Introduction to Neural Machine Translation

ATA59

Carola F. Berger


TURN OFF
2-WAY RADIO
AND
CELL PHONE

Disclaimer 1 – This Presentation

Disclaimer 1 – This Presentation

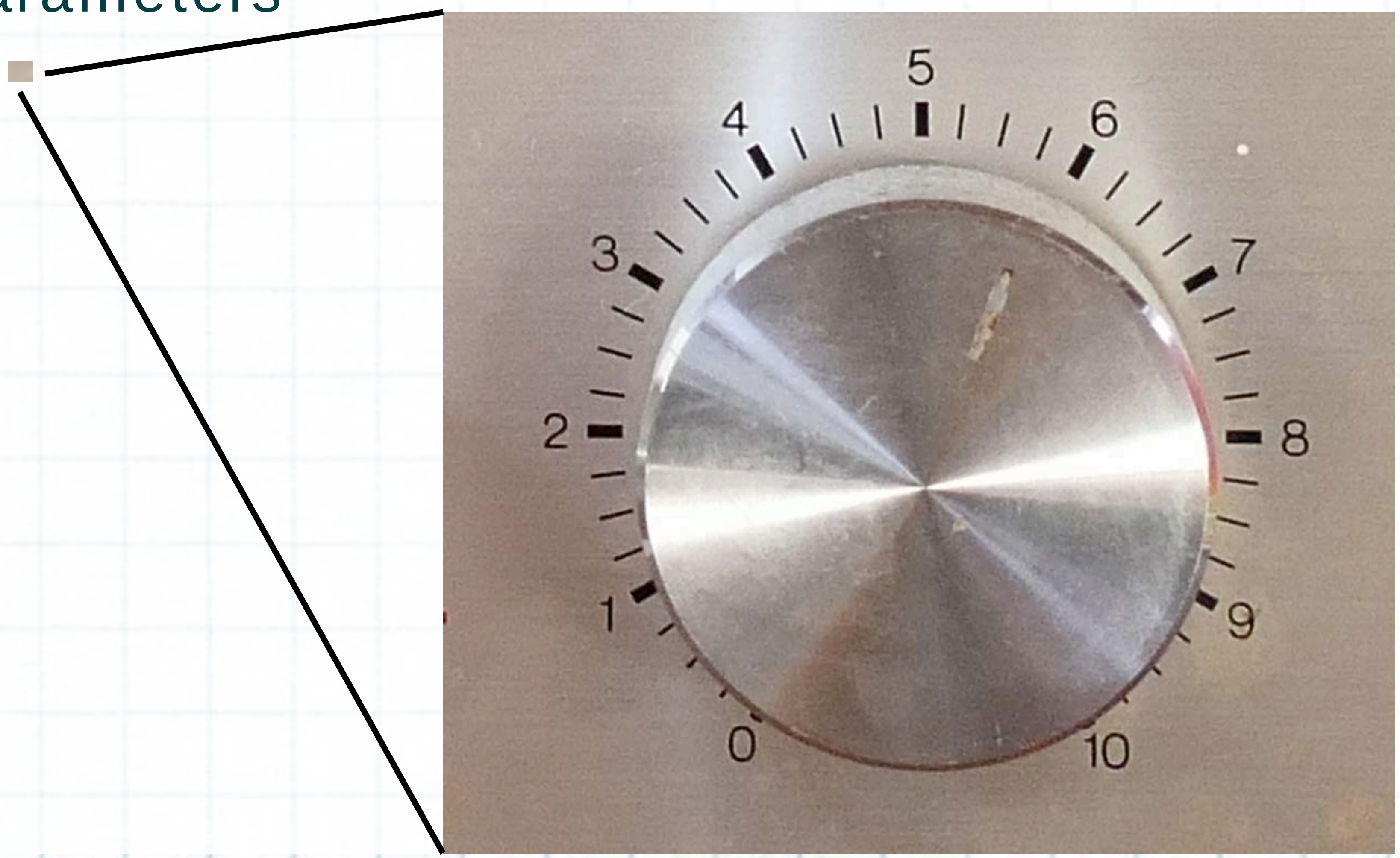
Disclaimer 2

After this presentation, you will hopefully understand how neural MT works, but you will not understand what it does.

Disclaimer 2

After this presentation, you will hopefully understand how neural MT works, but you will not understand what it does.

~100 million parameters



Outline

- Brief recap: previous MT approaches
- How do neural networks work?
- How do words get into and out of a neural network?

MT Approaches

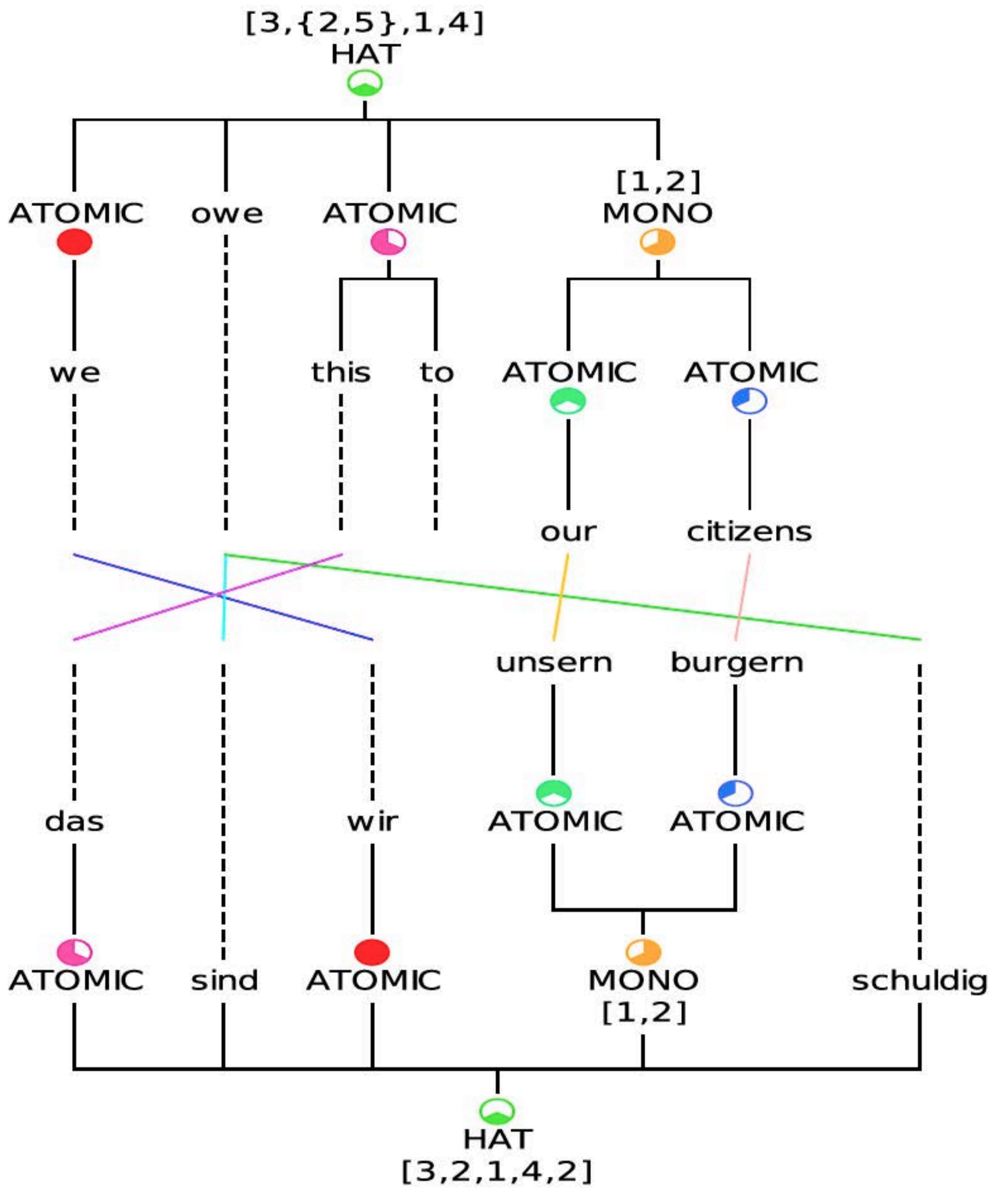
- Rules based:
  Basically just grammar + dictionary
- Statistical MT:
  Chop sentences up into n-grams (sequences of n words) or phrases.
  Training of the engine: Calculate frequency = probability in source and target language.
  Translation after training: Chop source sentences into n-grams or phrases, apply previously calculated probabilities.
  1-gram SMT = ?

MT Approaches

- Rules based:
  Basically just grammar + dictionary
- Statistical MT:
  Chop sentences up into n-grams (sequences of n words) or phrases.
  Training of the engine: Calculate frequency = probability in source and target language.
  Translation after training: Chop source sentences into n-grams or phrases, apply previously calculated probabilities.
  1-gram SMT = dictionary

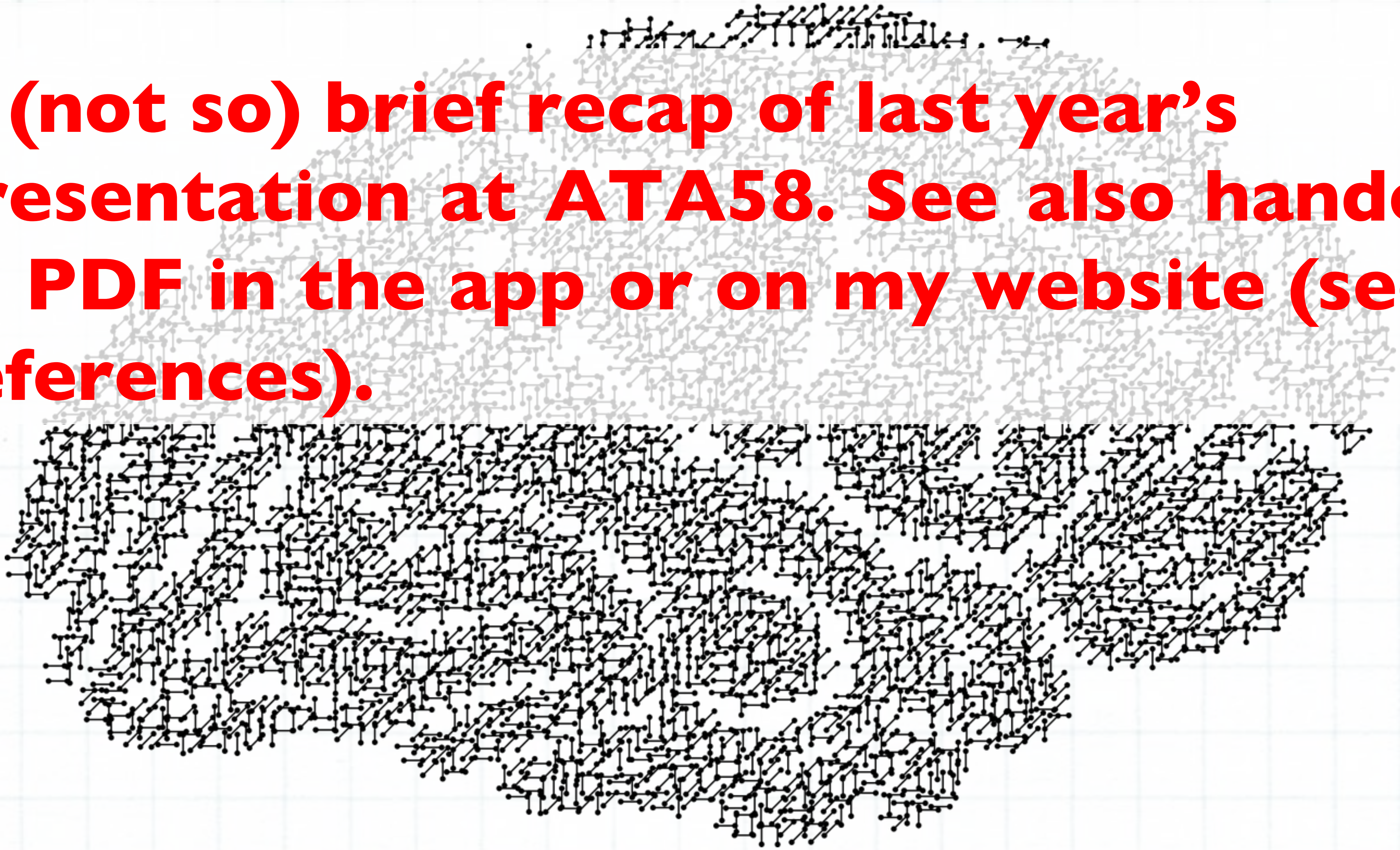
  Pros: Output is deterministic. No words missing.
  Cons: Context!
- Neural MT – this presentation

Statistical MT



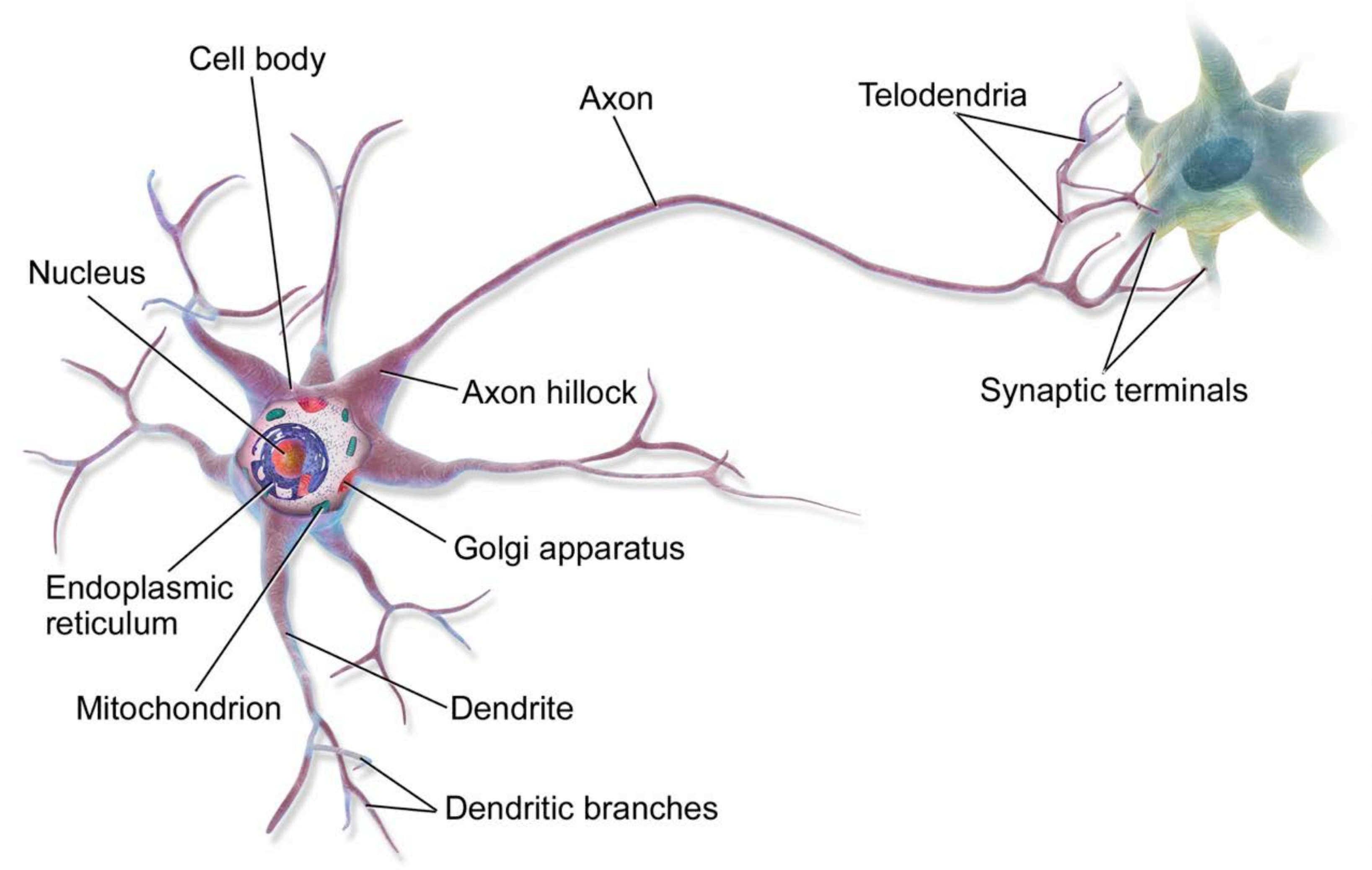
From G. M. de Buy Wenninger, K. Sima'an, PBML No. 101, April 2014, pp. 43

How Do Neural Networks Work?



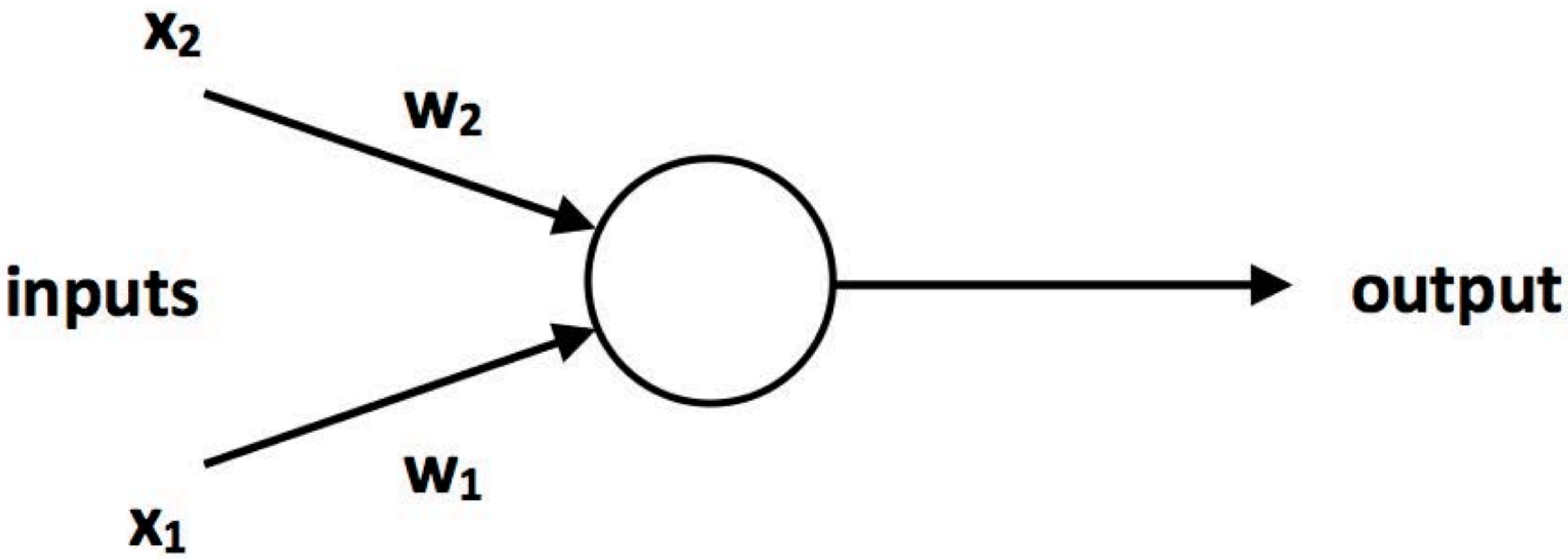
A (not so) brief recap of last year's presentation at ATA58. See also handouts as PDF in the app or on my website (see references).

Biological Neuron

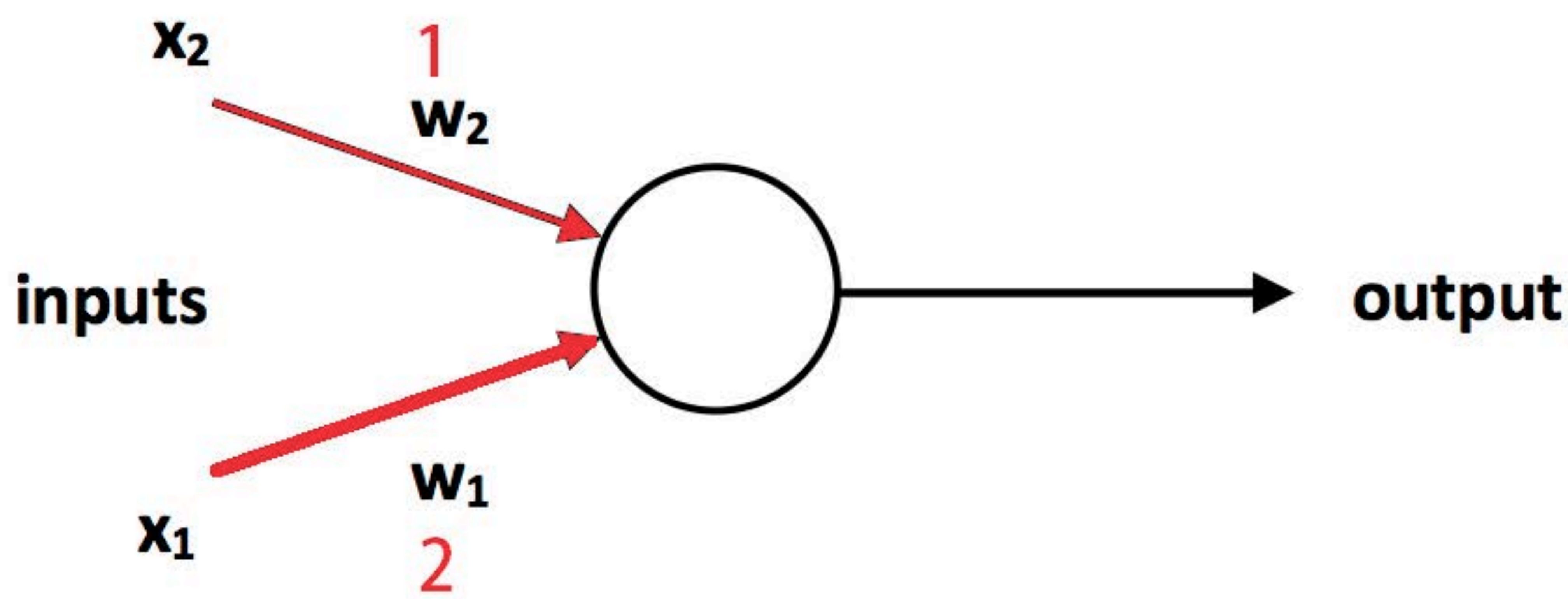


Bruce Blaus, https://commons.wikimedia.org/wiki/File:Blausen_0657_MultipolarNeuron.png

Artificial Neuron

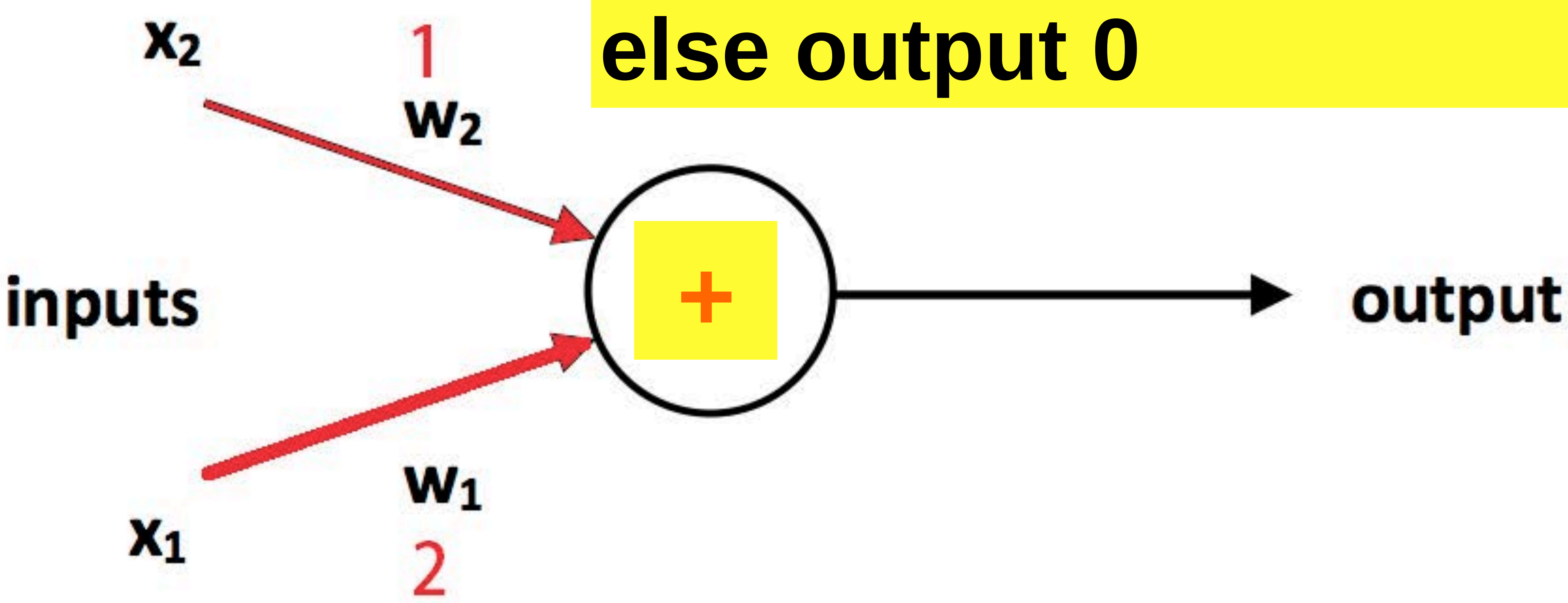


Artificial Neuron



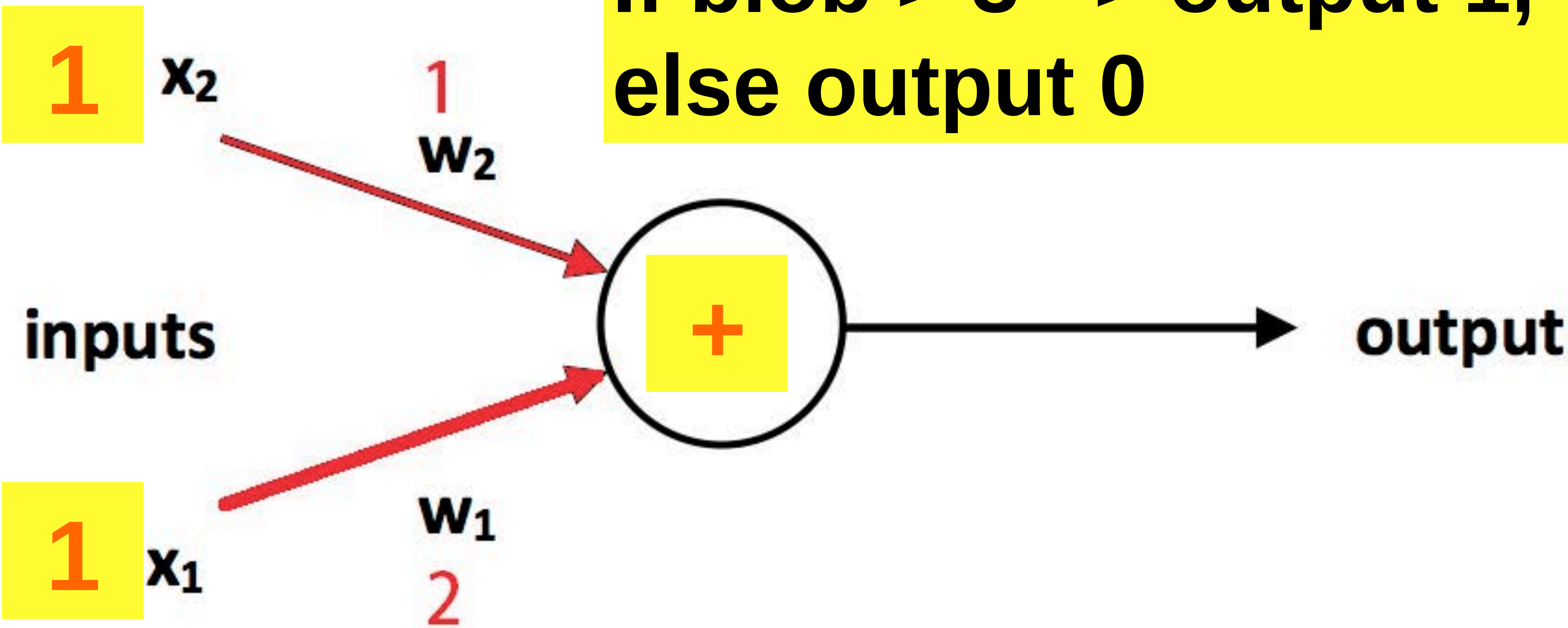
Artificial Neuron - Perceptron

If blob > 3 => output 1,
else output 0

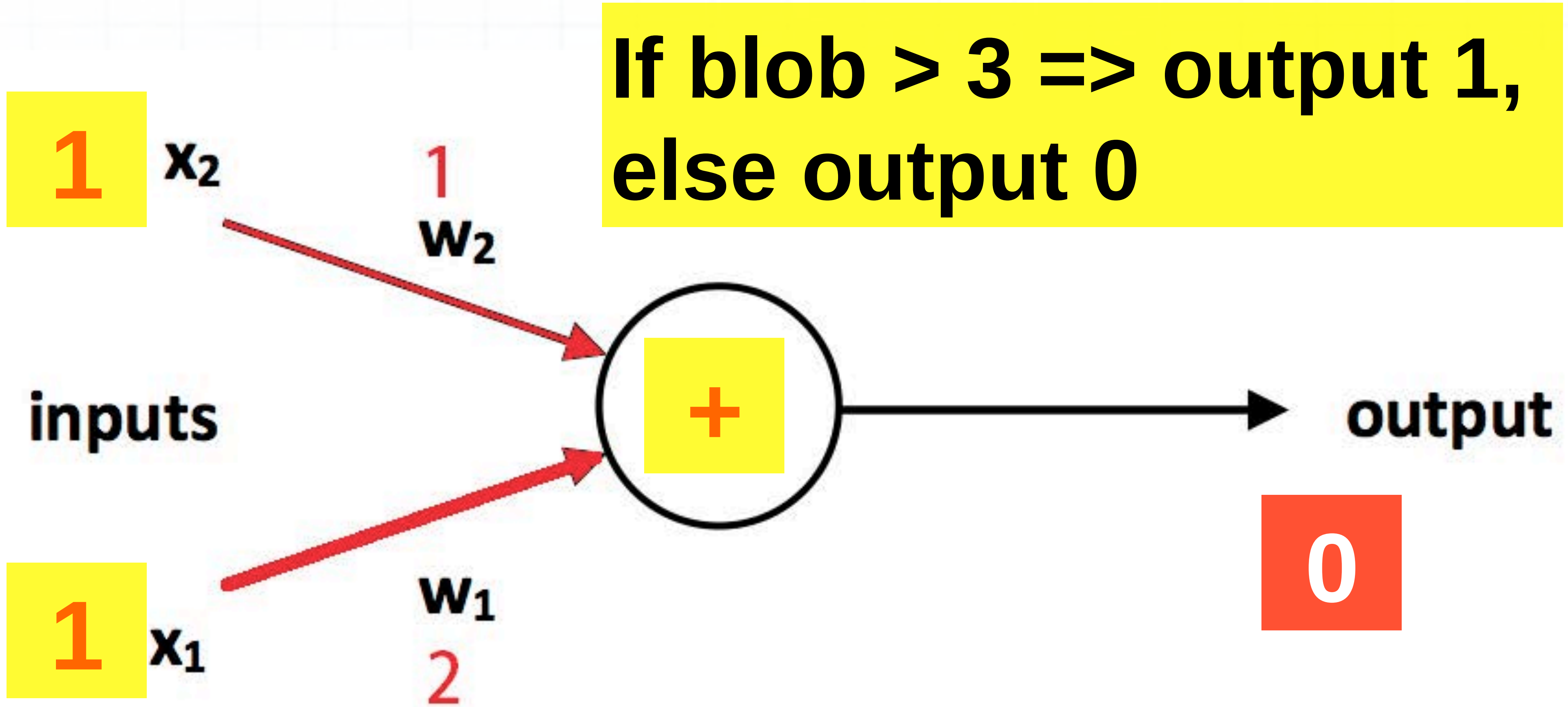


Artificial Neuron - Perceptron

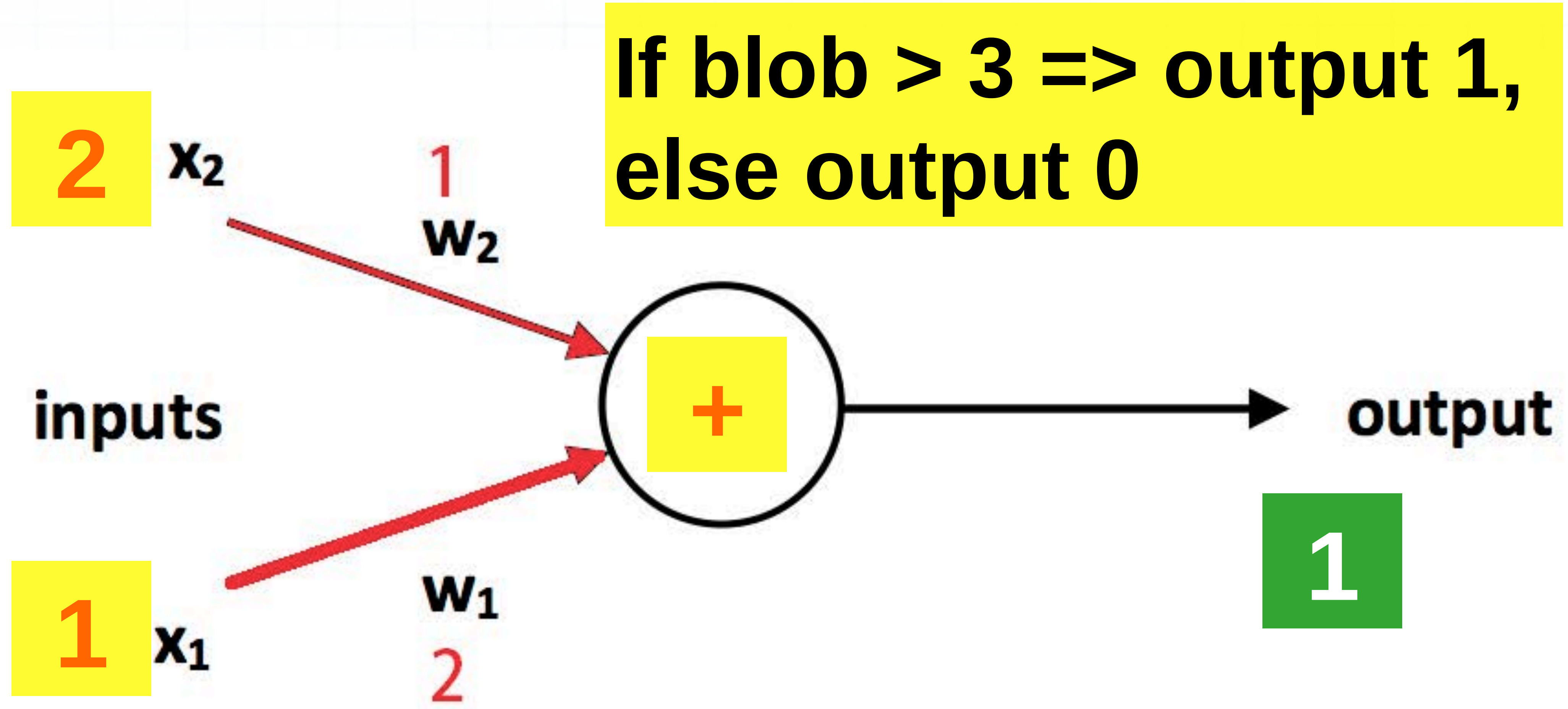
If blob > 3 => output 1,
else output 0



Artificial Neuron - Perceptron

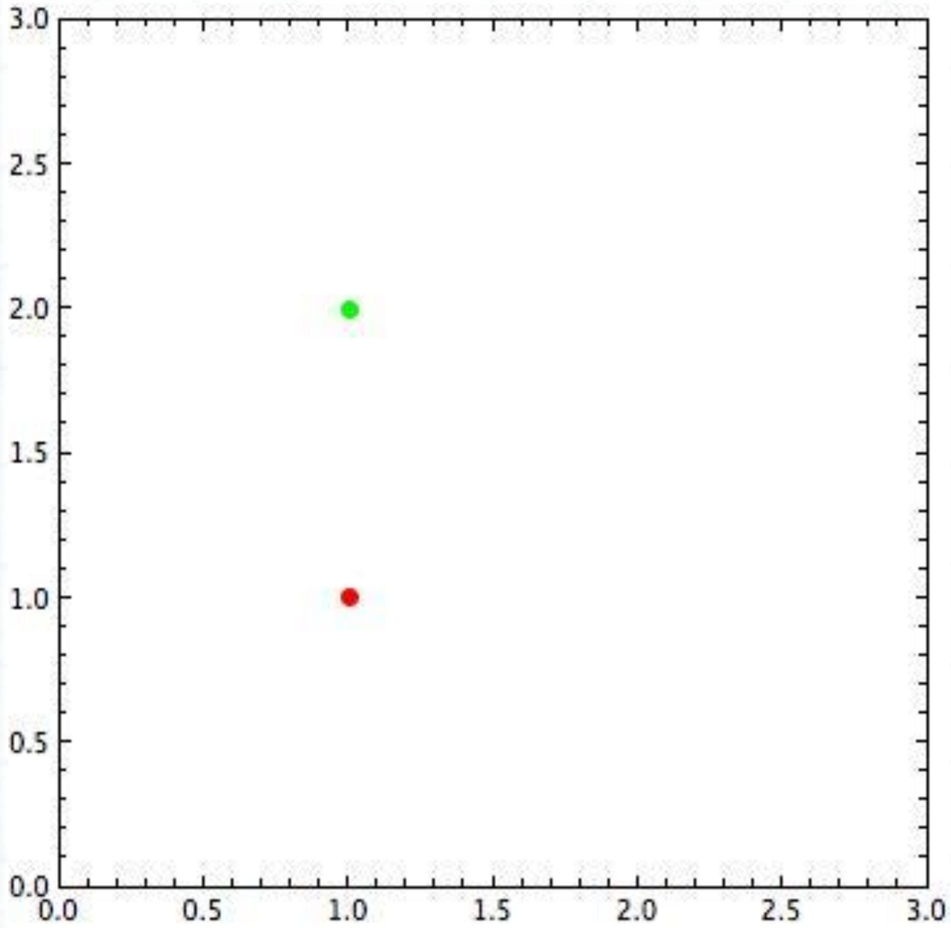


Artificial Neuron - Perceptron

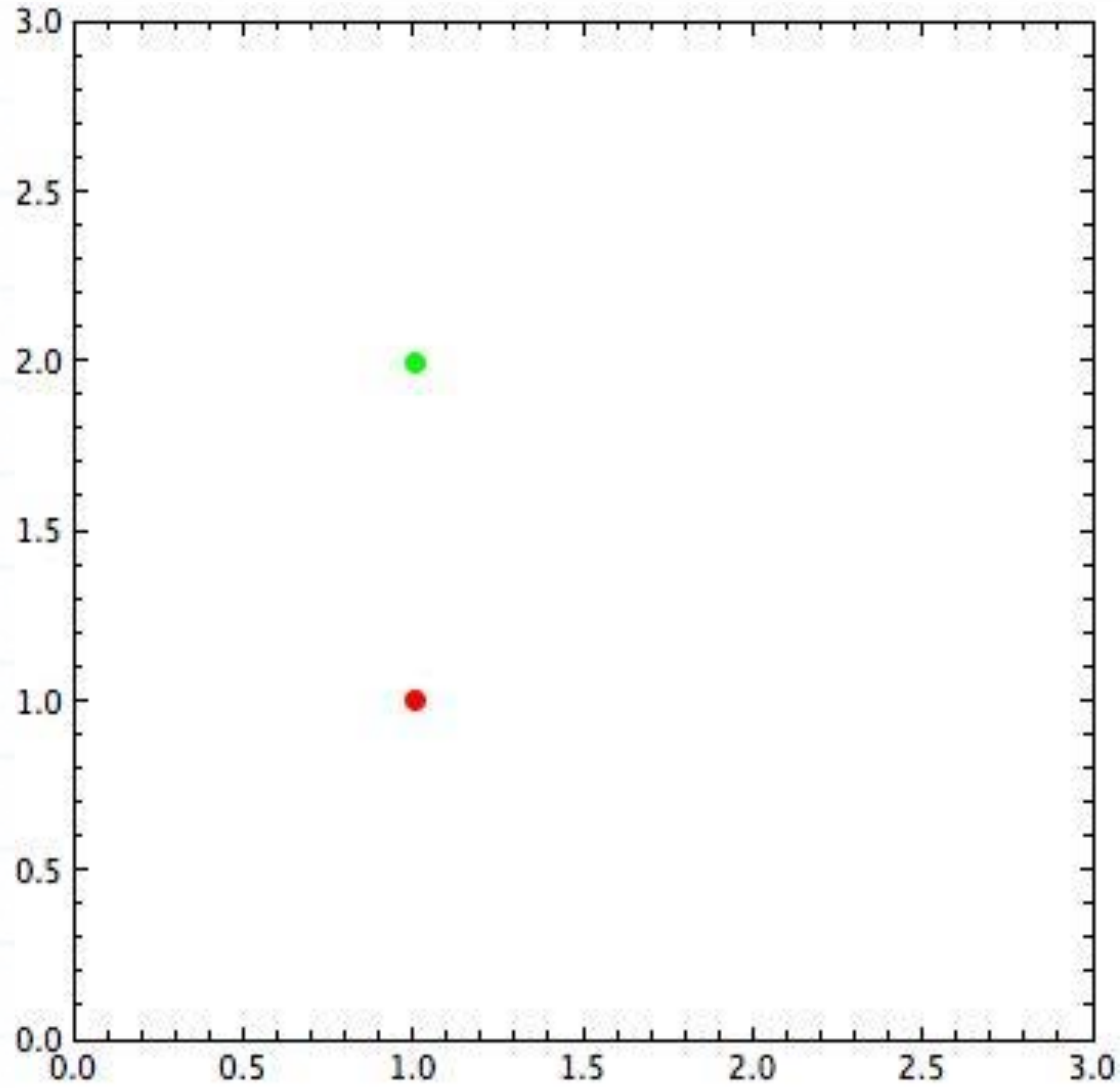


Artificial Neuron - Perceptron

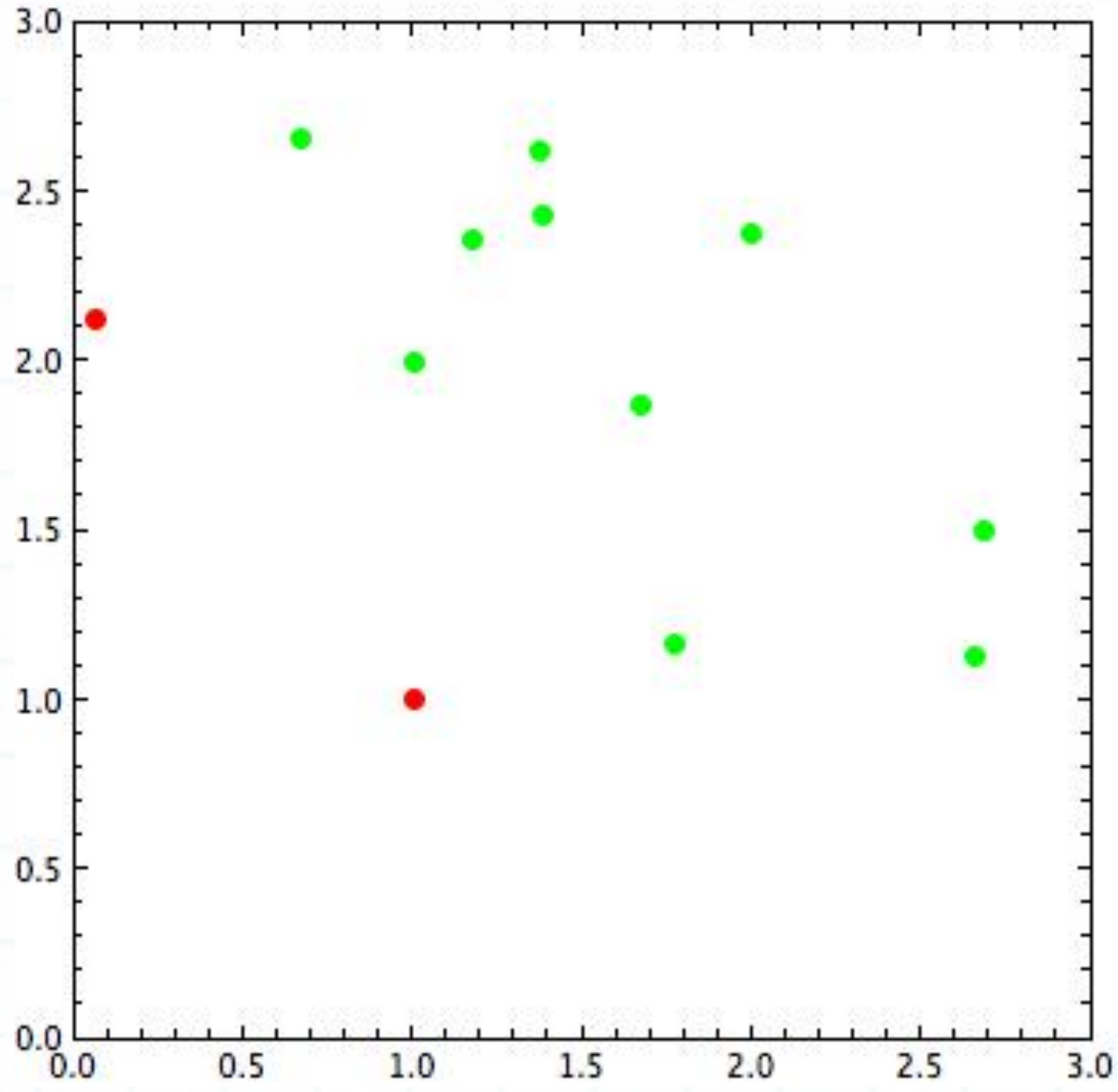
x1	x2	OUTPUT
1	1	0
1	2	1



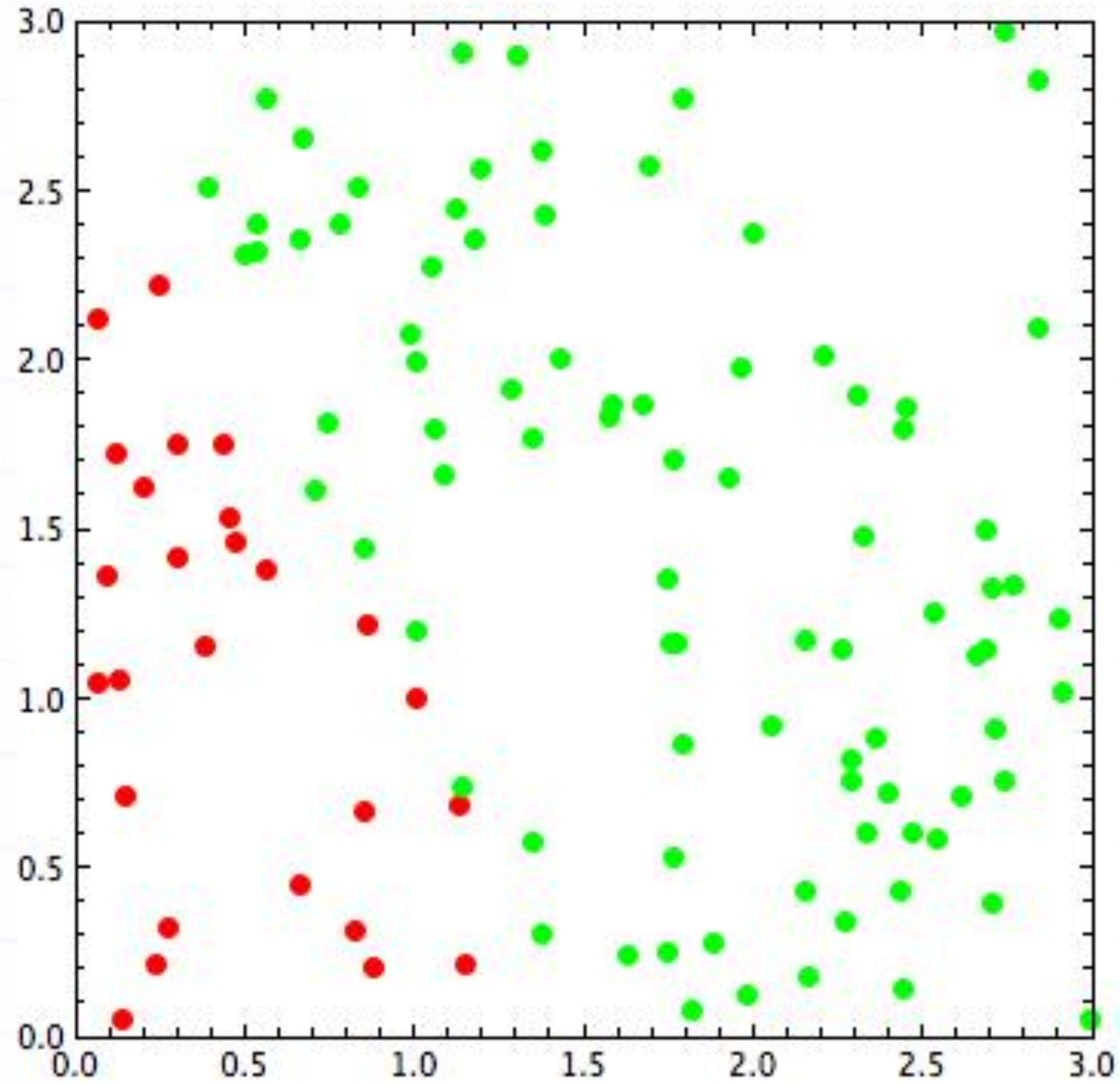
Artificial Neuron - Perceptron



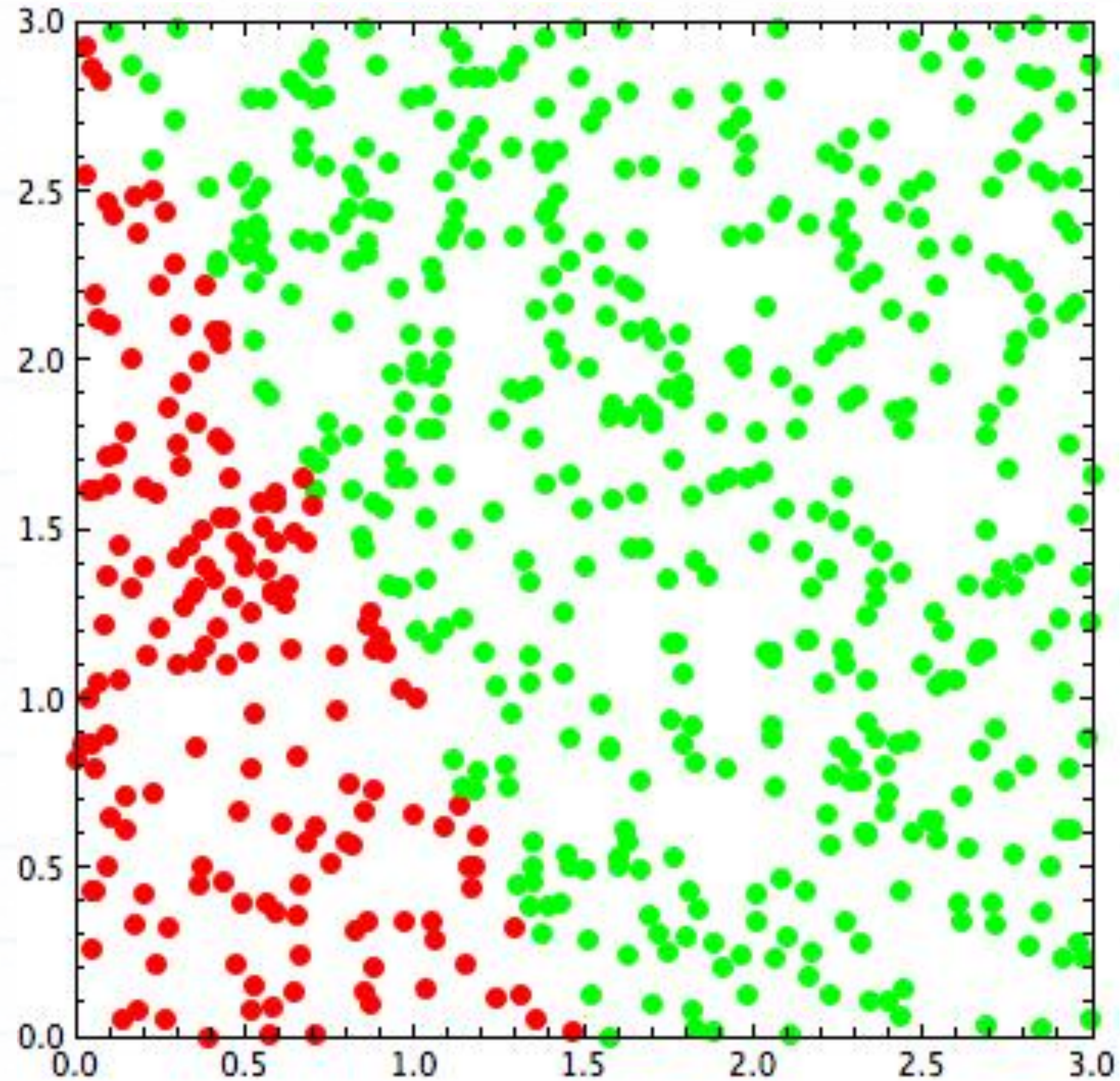
Artificial Neuron - Perceptron



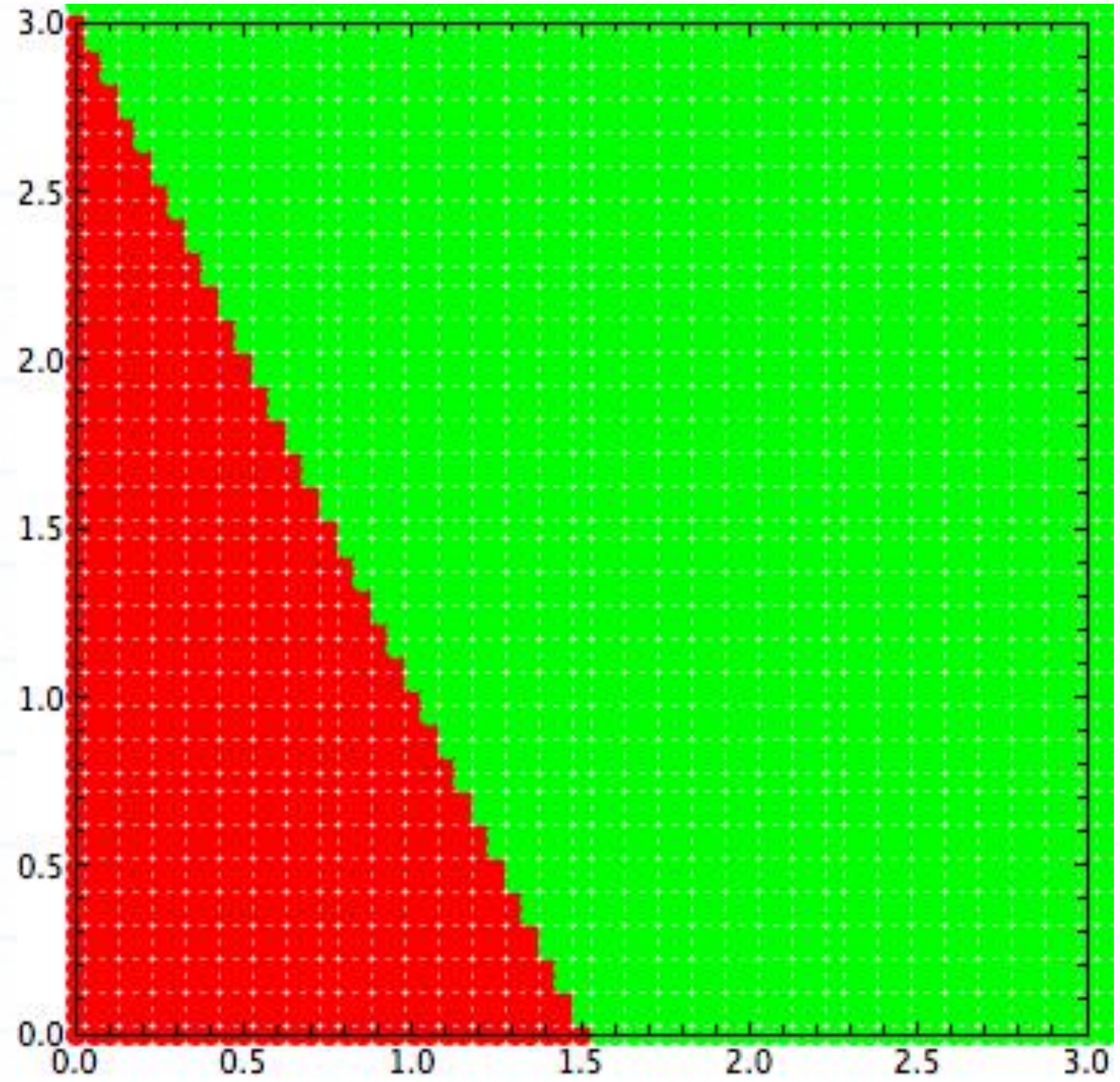
Artificial Neuron - Perceptron



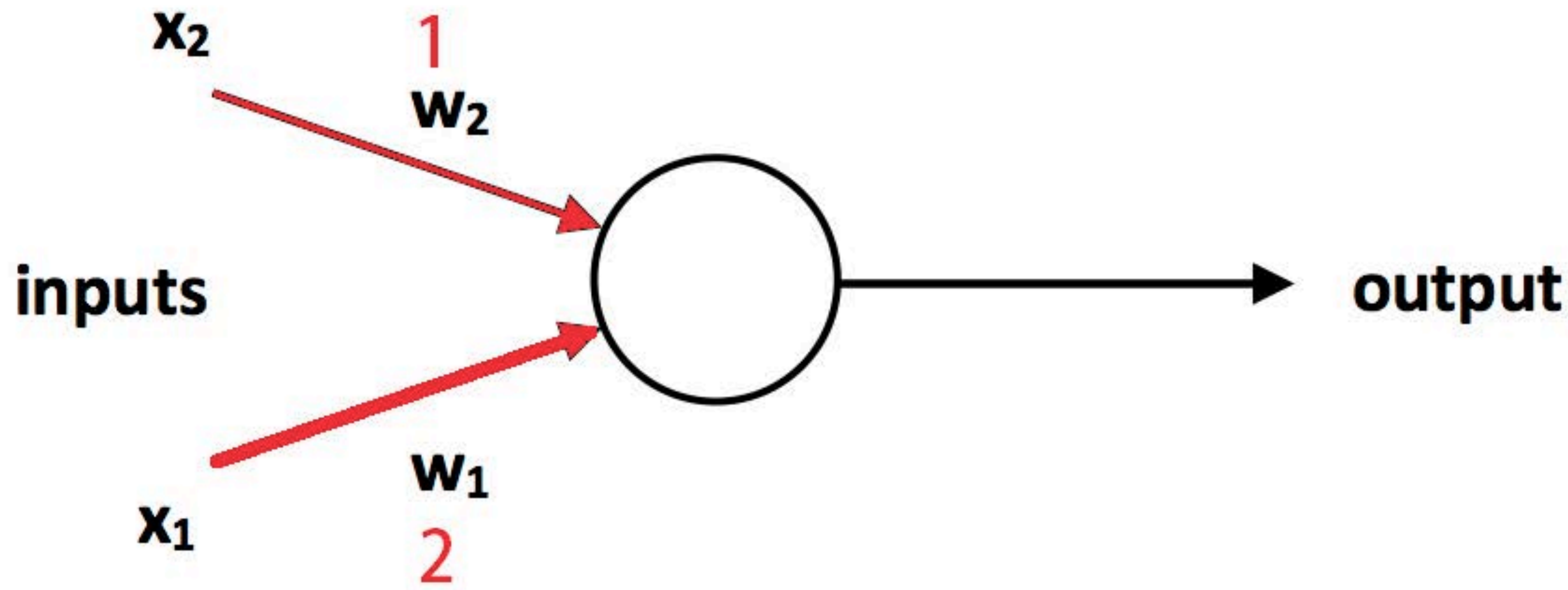
Artificial Neuron - Perceptron



Artificial Neuron - Perceptron

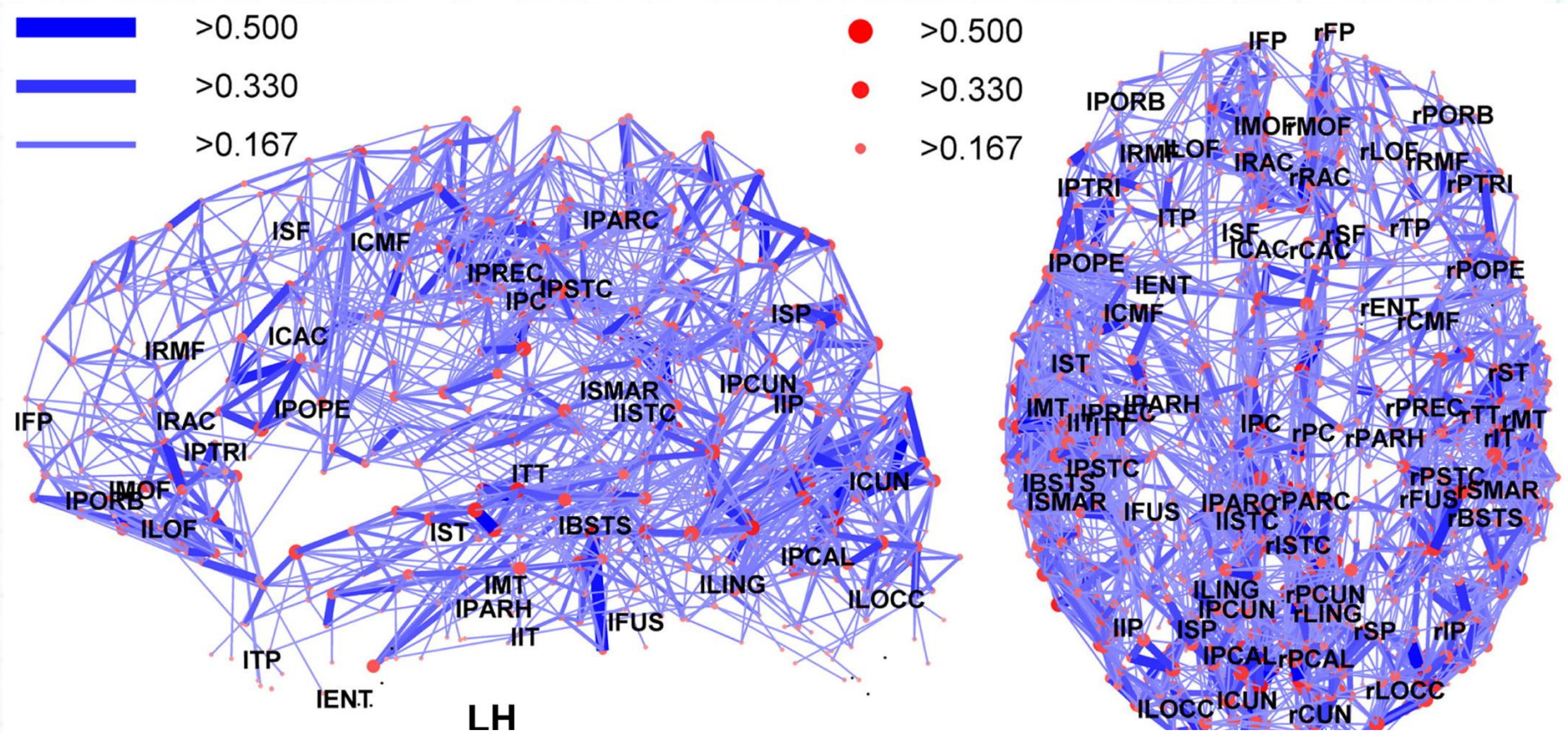


Artificial Neuron



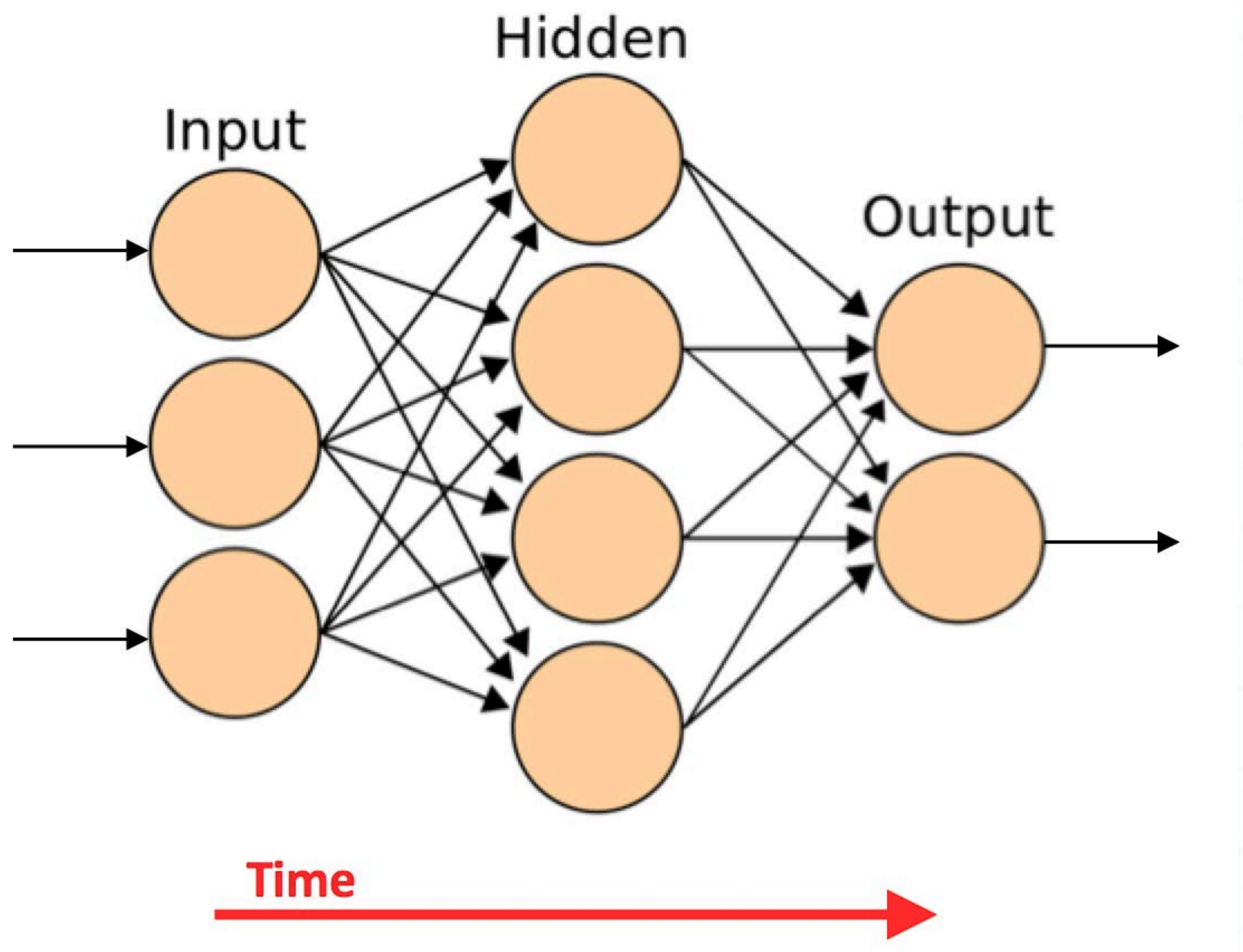
Neural Network

- Human brain:

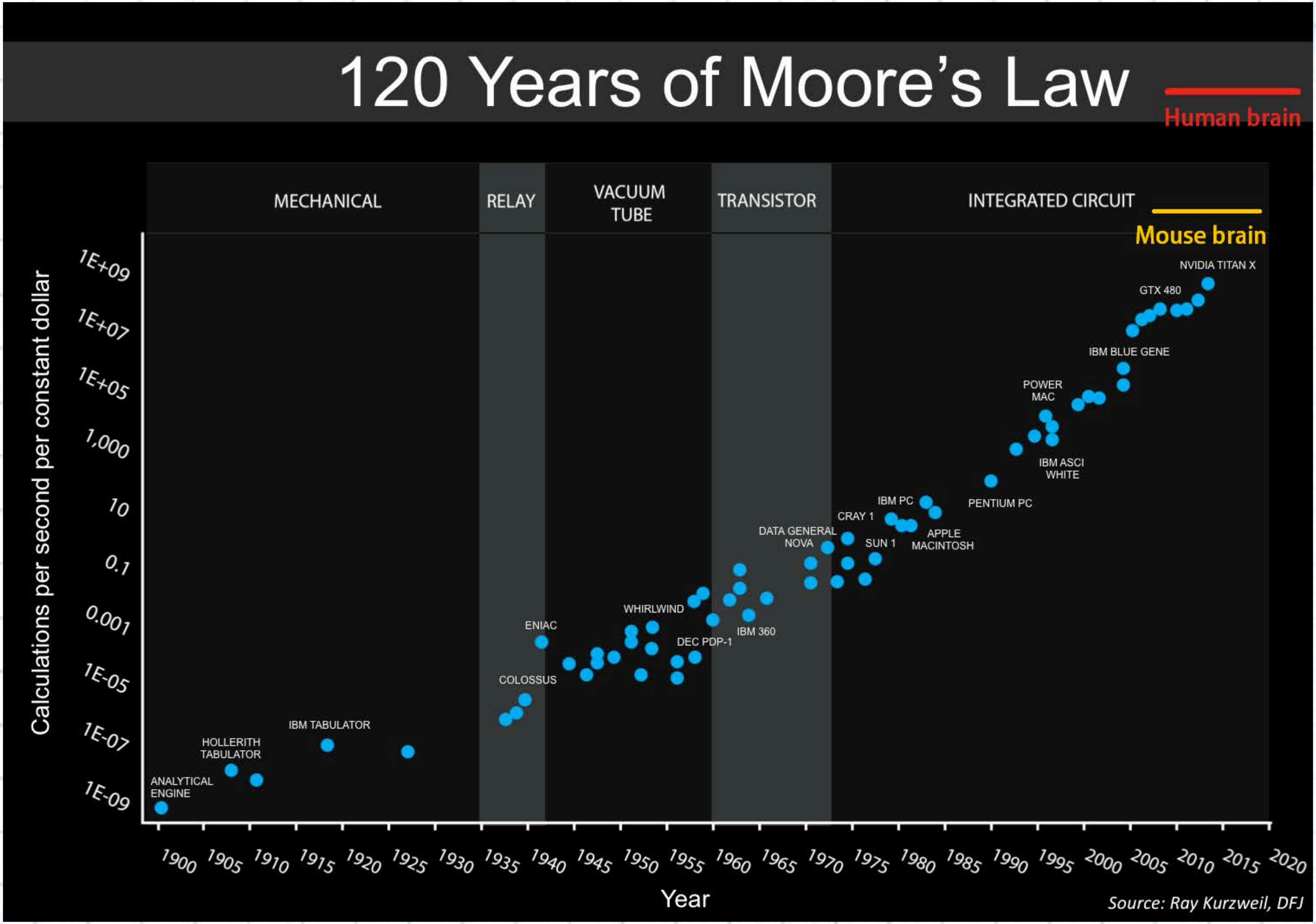


Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008) Mapping the structural core of human cerebral cortex. PLoS Biology Vol. 6, No. 7, e159.

Artificial Neural Network



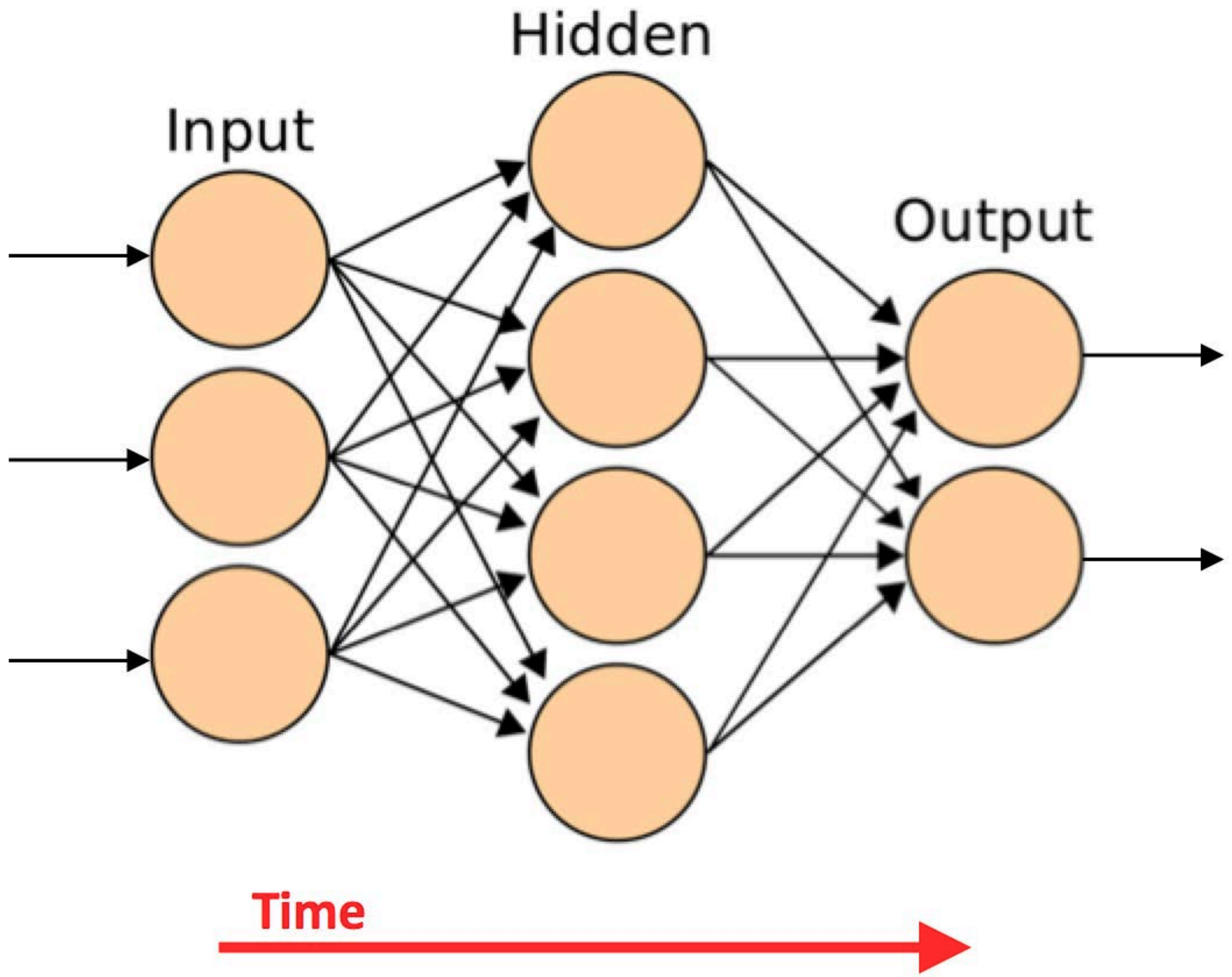
Adapted from: Cburnett, https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg



Adapted from: S. Jurvetson, https://www.flickr.com/photos/jurvetson/31409423572/

Artificial Neural Network

- Training:

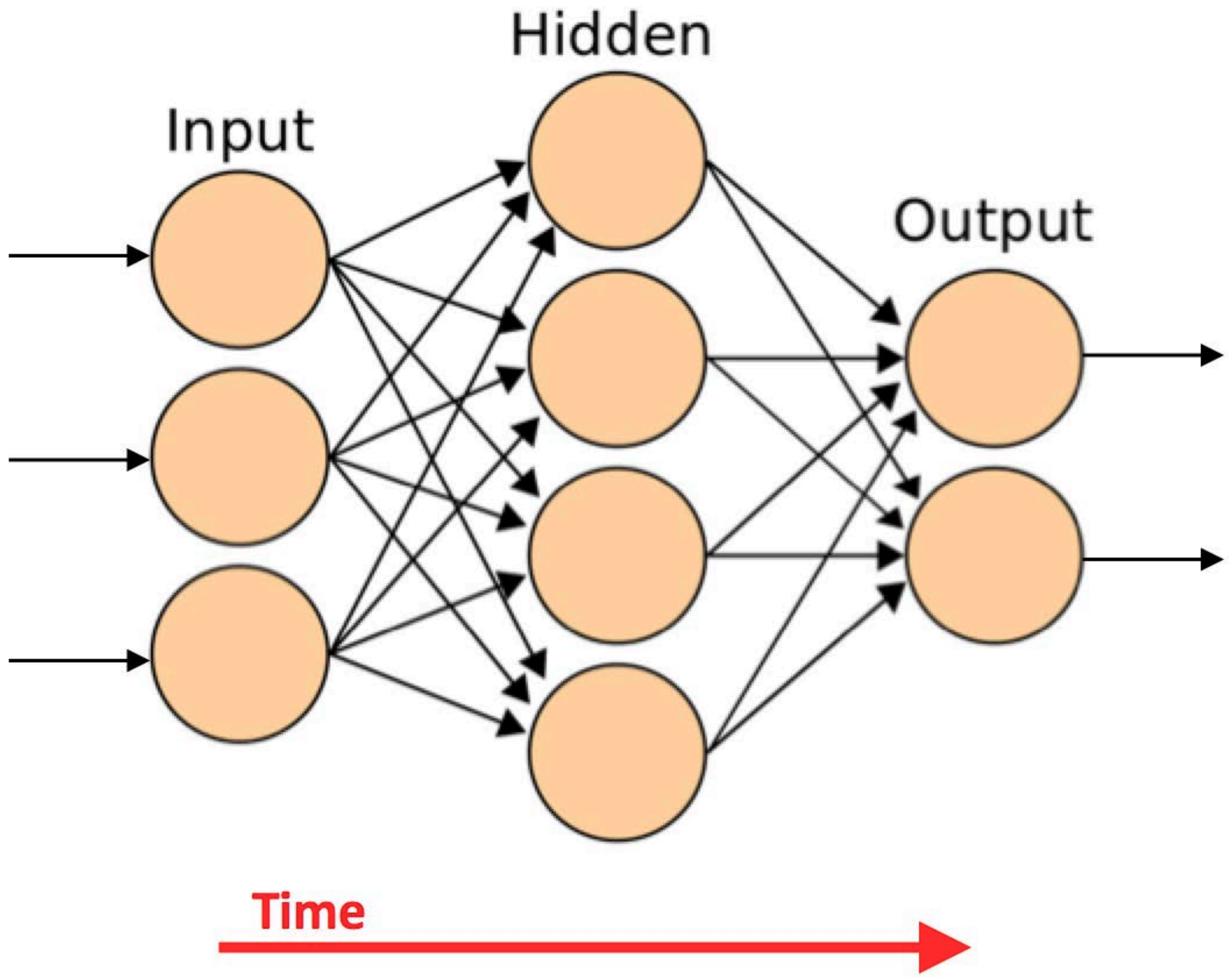


Adapted from: Cburnett, https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

Artificial Neural Network

- Training:

Feed in training data

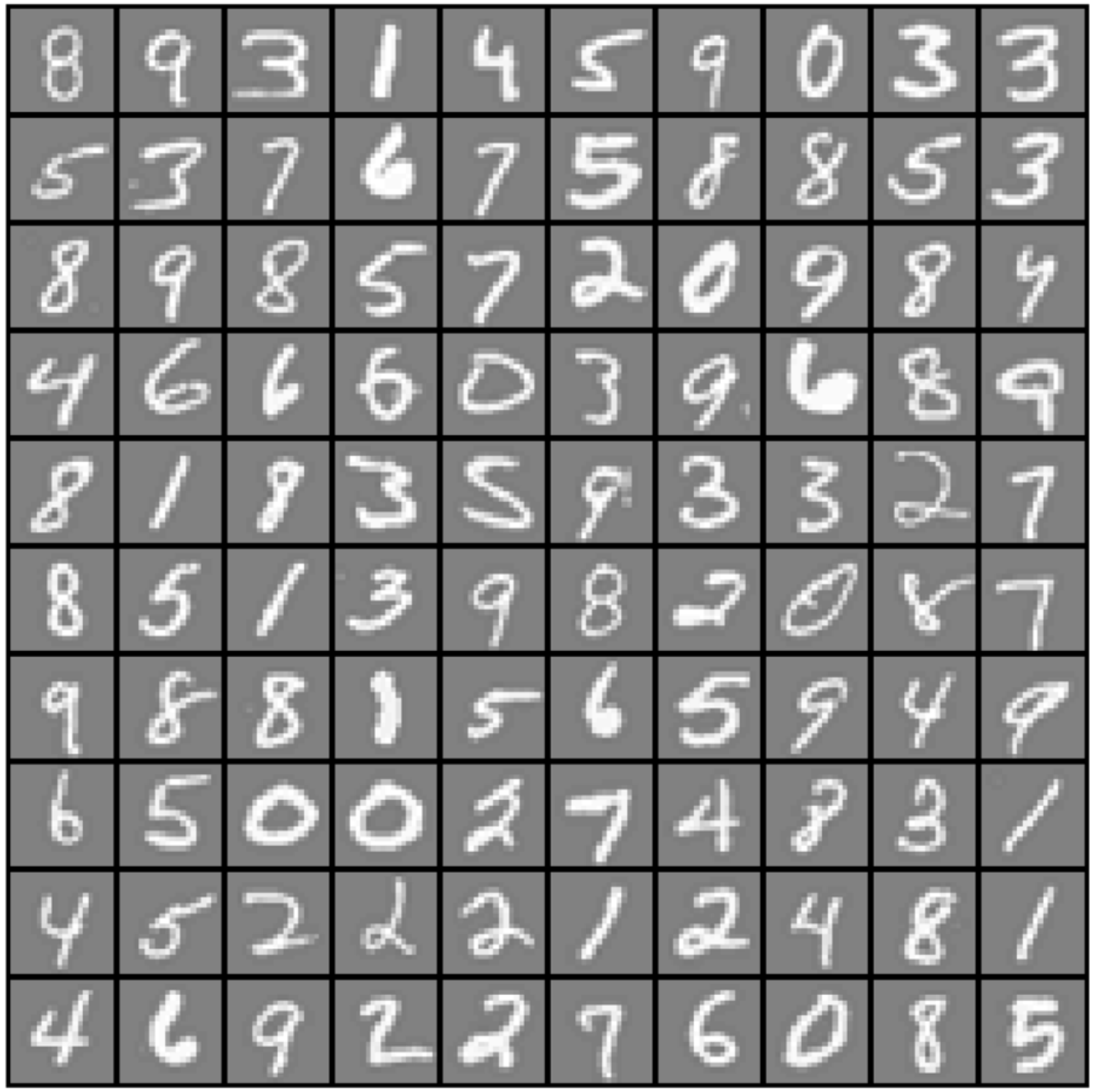


Adapt weights ("arrows") according to difference between desired output and actual output, e.g. by backpropagation

Adapted from: Cburnett, https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

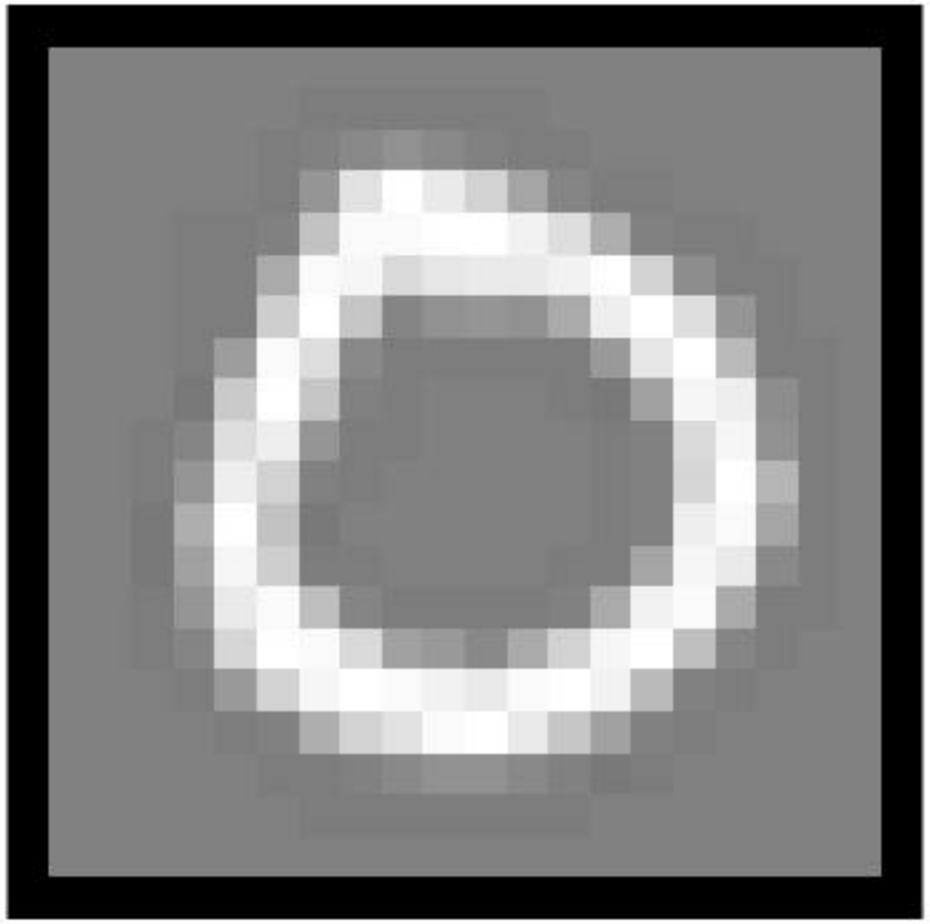
Neural Net Example – Digit Recognition

Sample input data



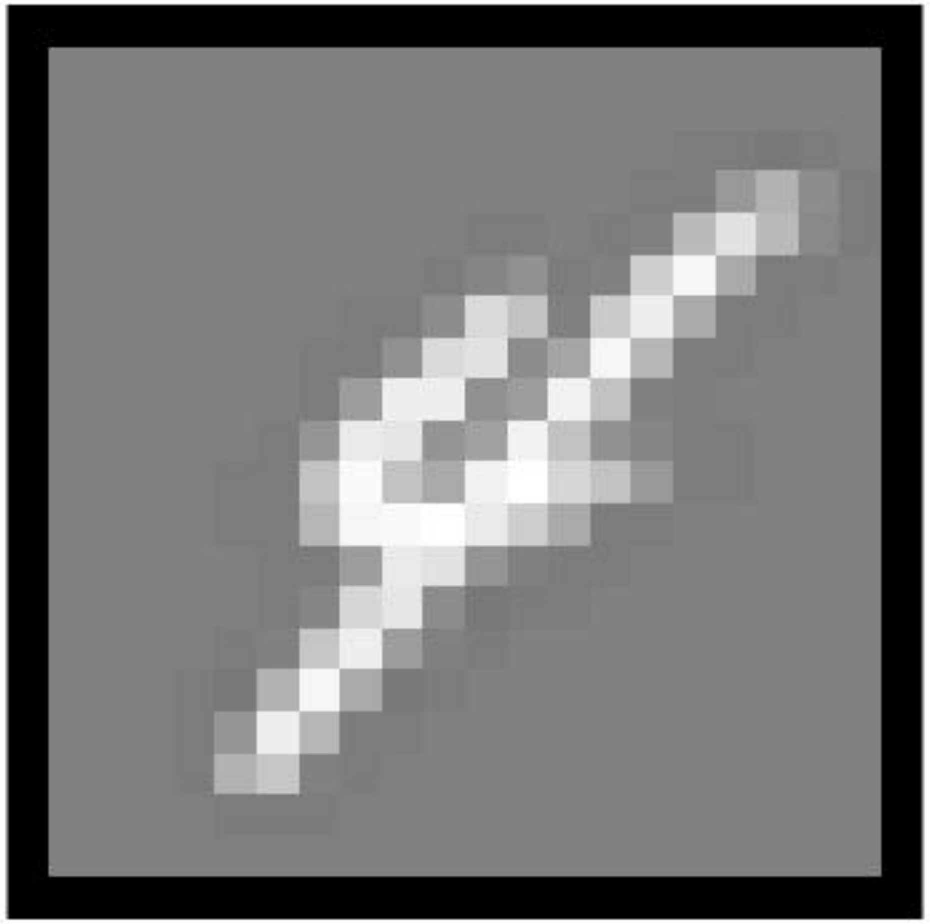
Neural Net Example – Digit Recognition

Sample input (20x20 pixels)



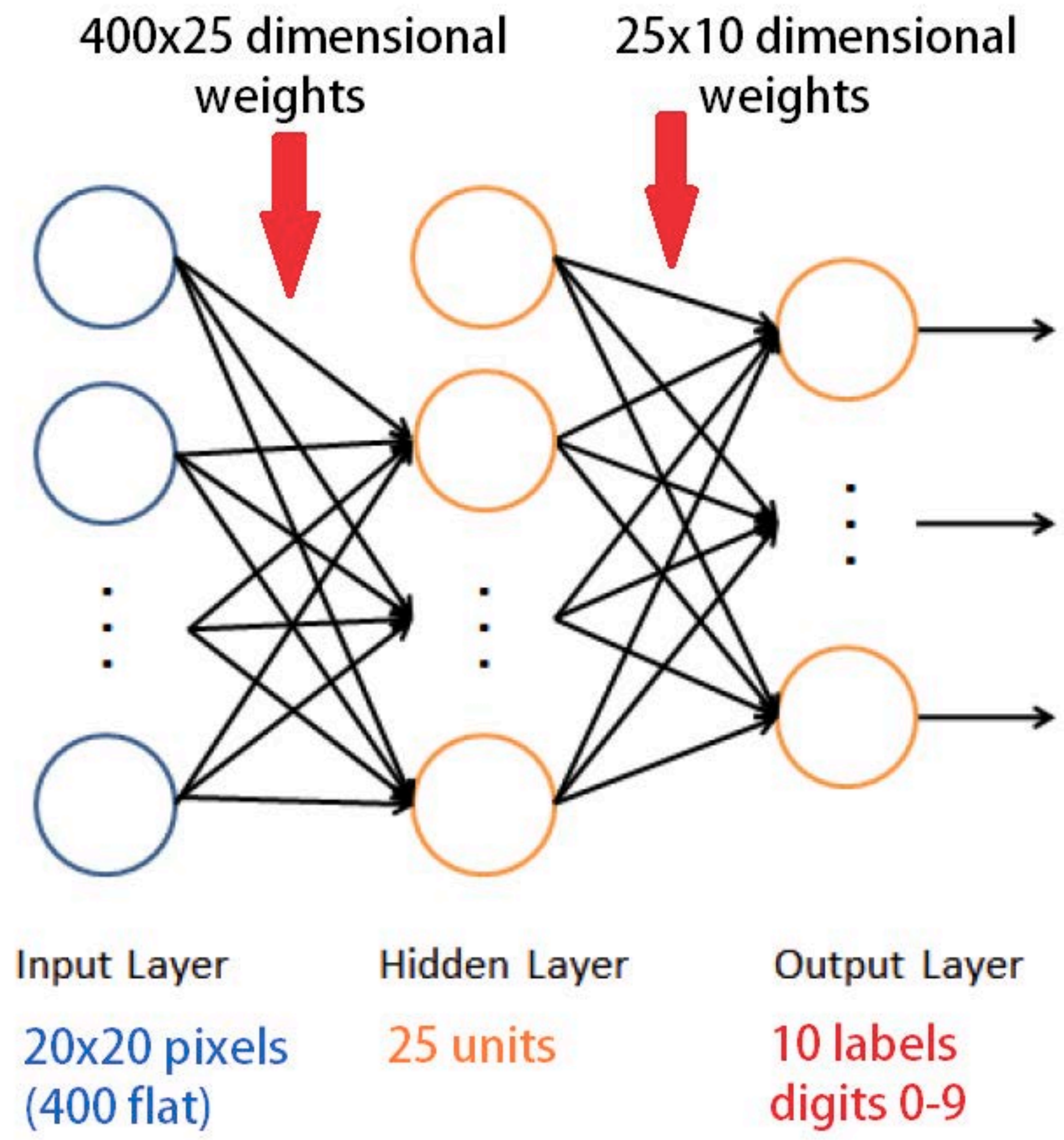
Neural Net Example – Digit Recognition

Sample input (20x20 pixels)



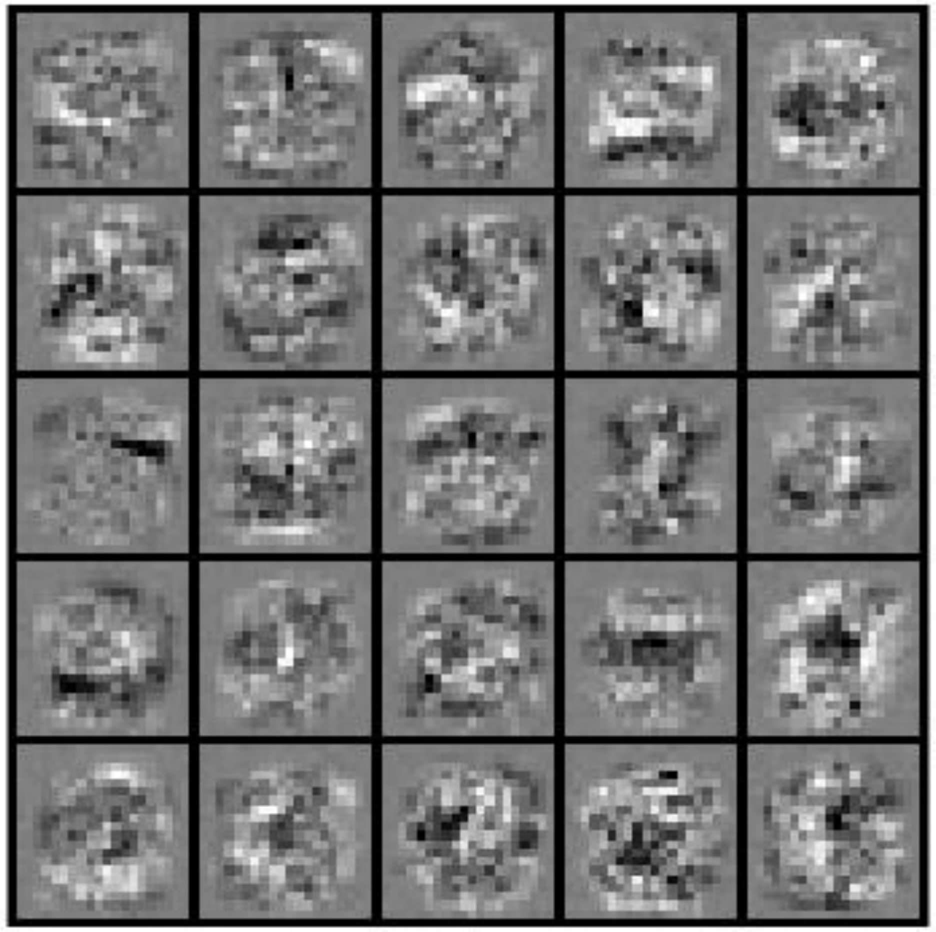
Neural Net Example – Digit Recognition

Weights



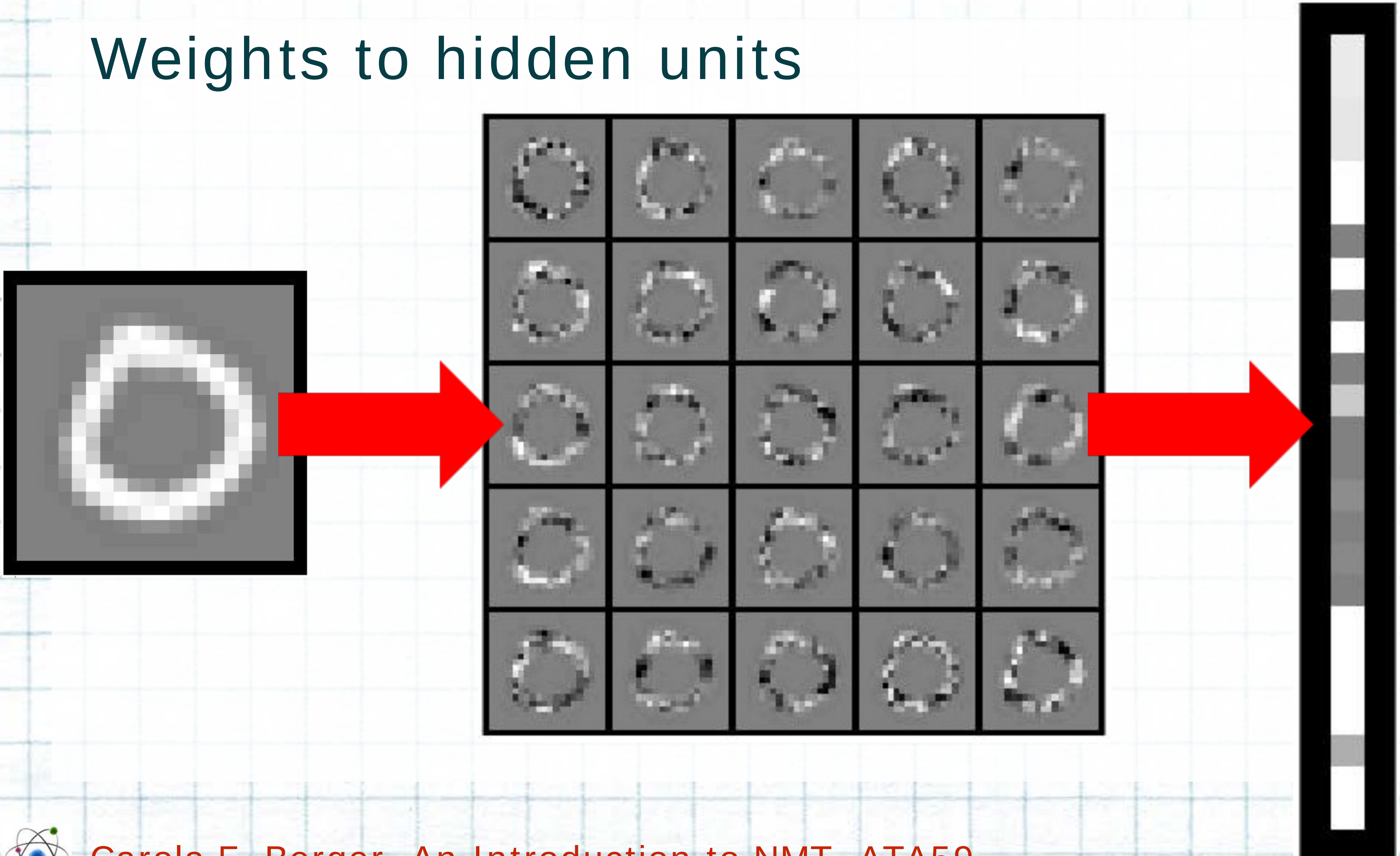
Neural Net Example – Digit Recognition

Weights to hidden units –
"feature" extraction



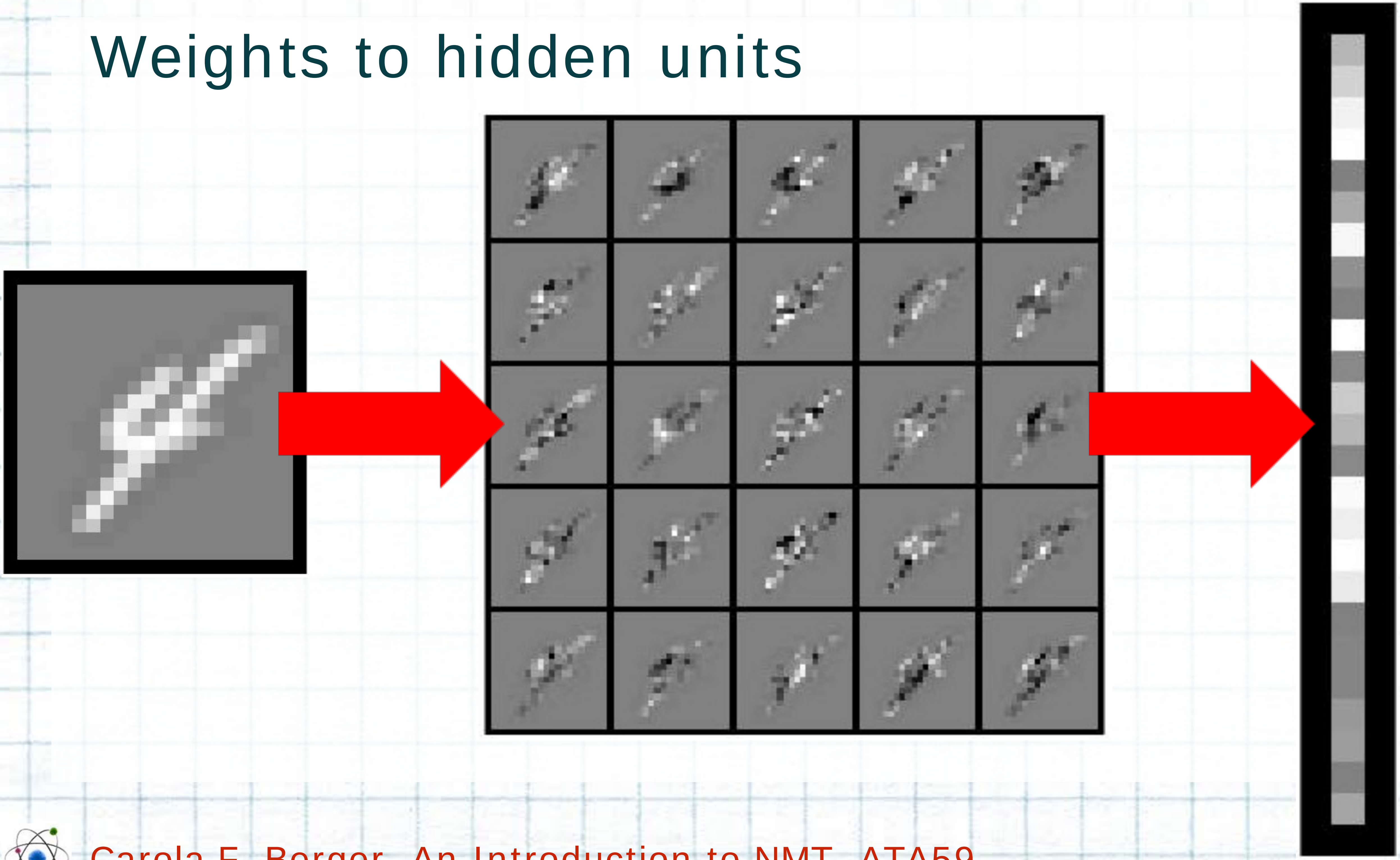
Neural Net Example – Digit Recognition

Weights to hidden units



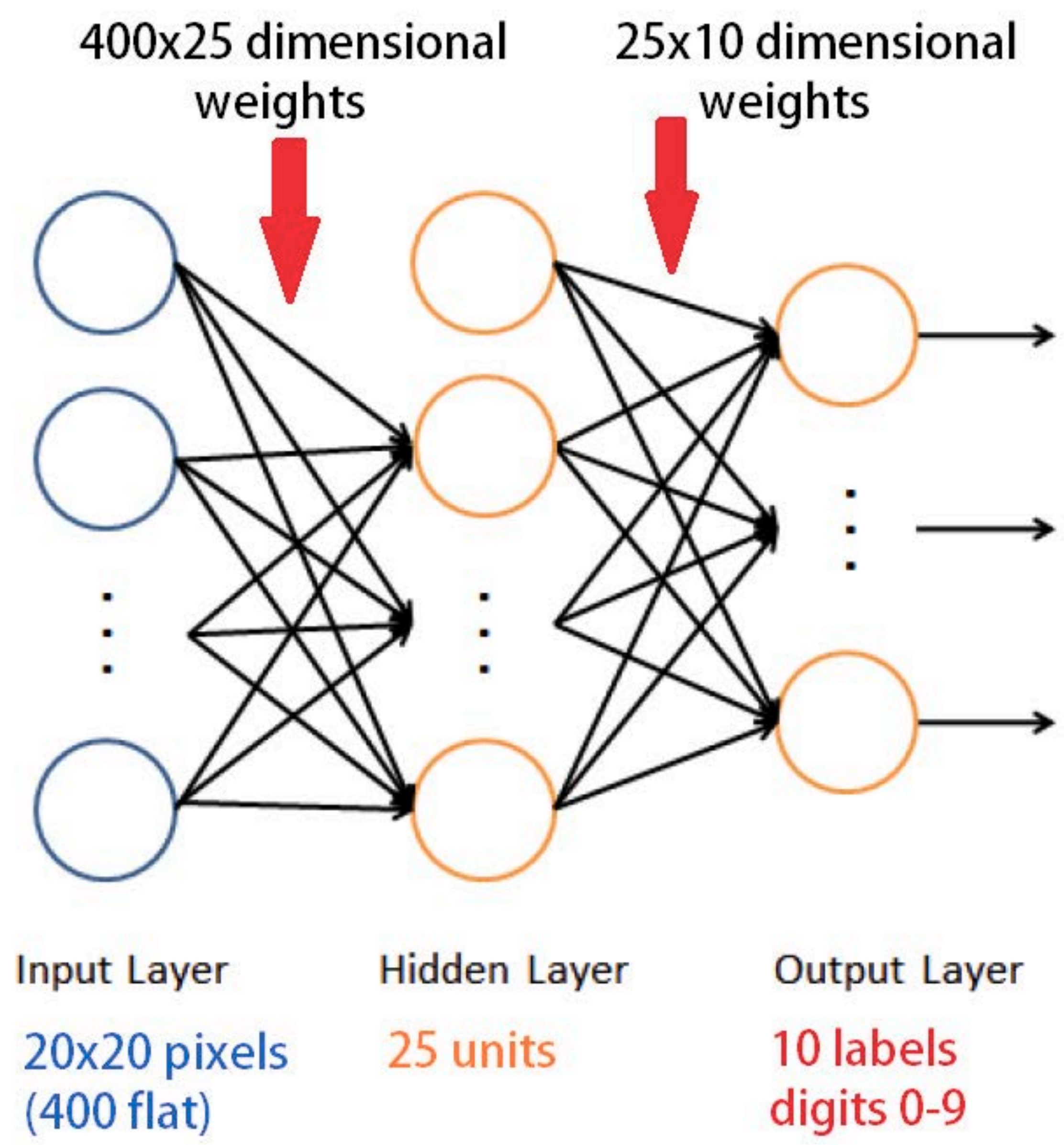
Neural Net Example – Digit Recognition

Weights to hidden units



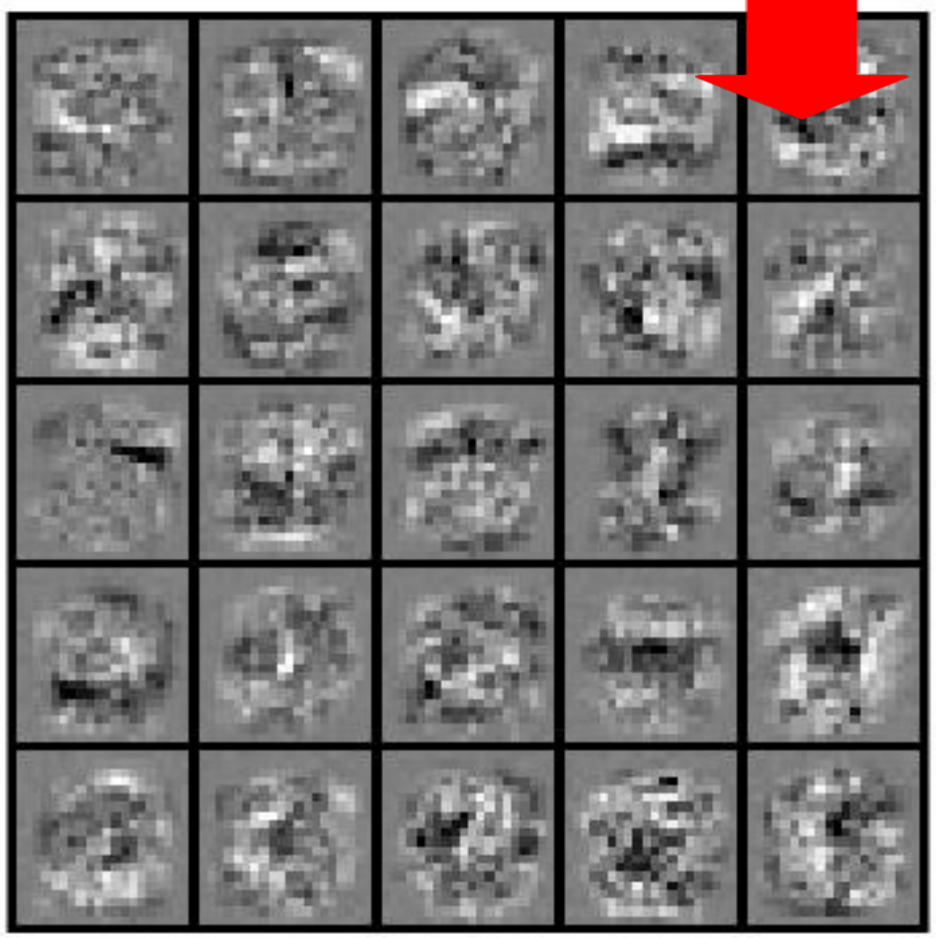
Neural Net Example – Digit Recognition

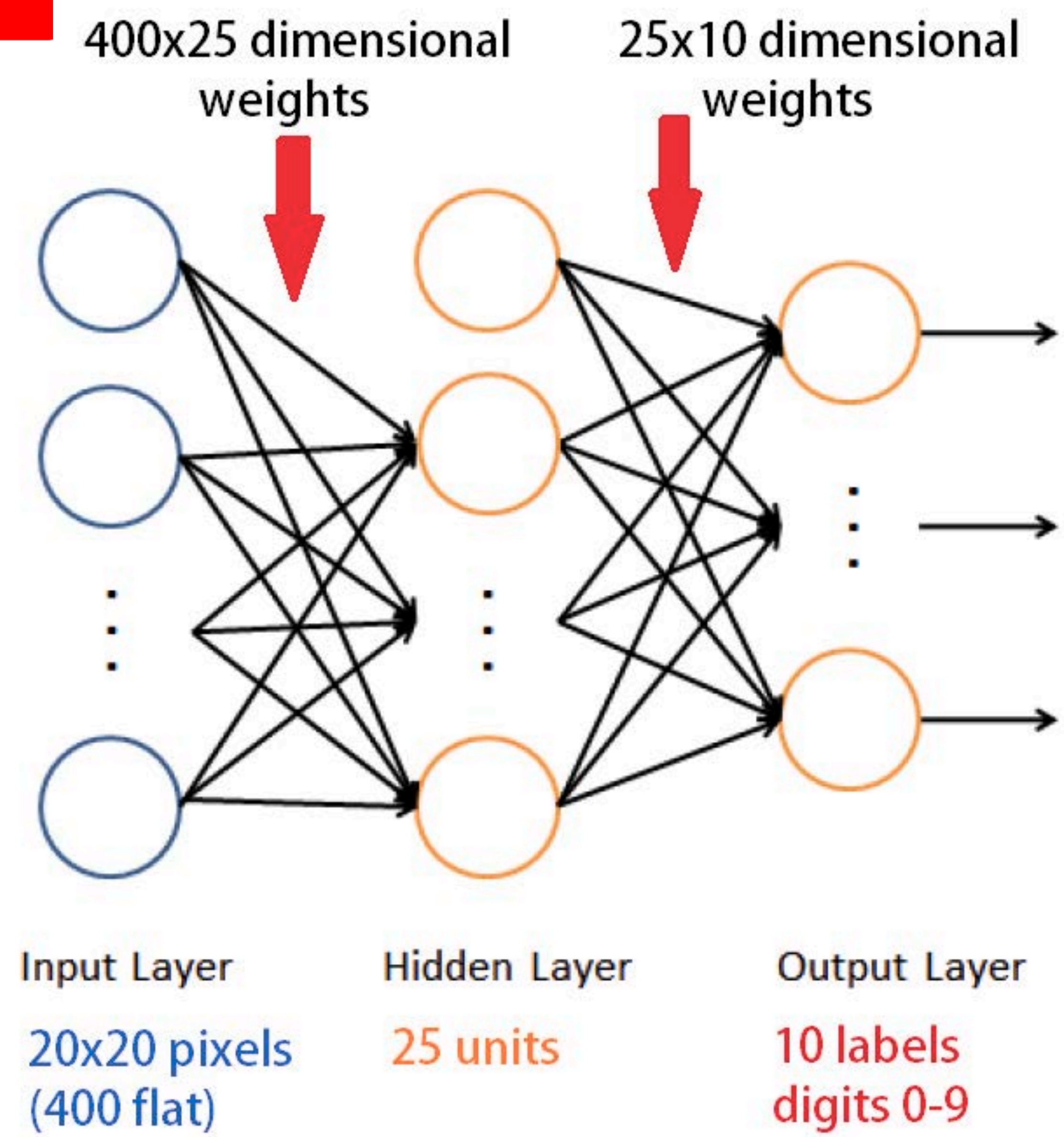
Hidden units to output

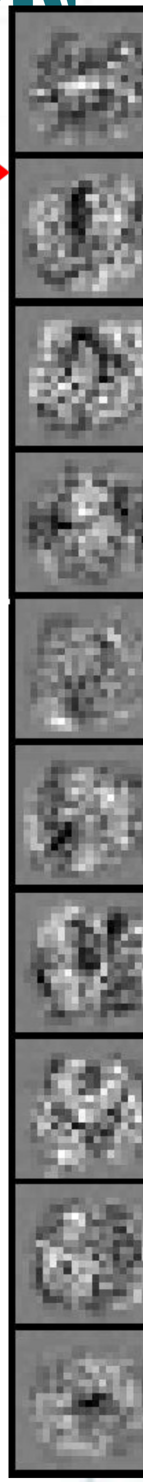


Neural Net Example – Digit Recognition

Hidden units to output

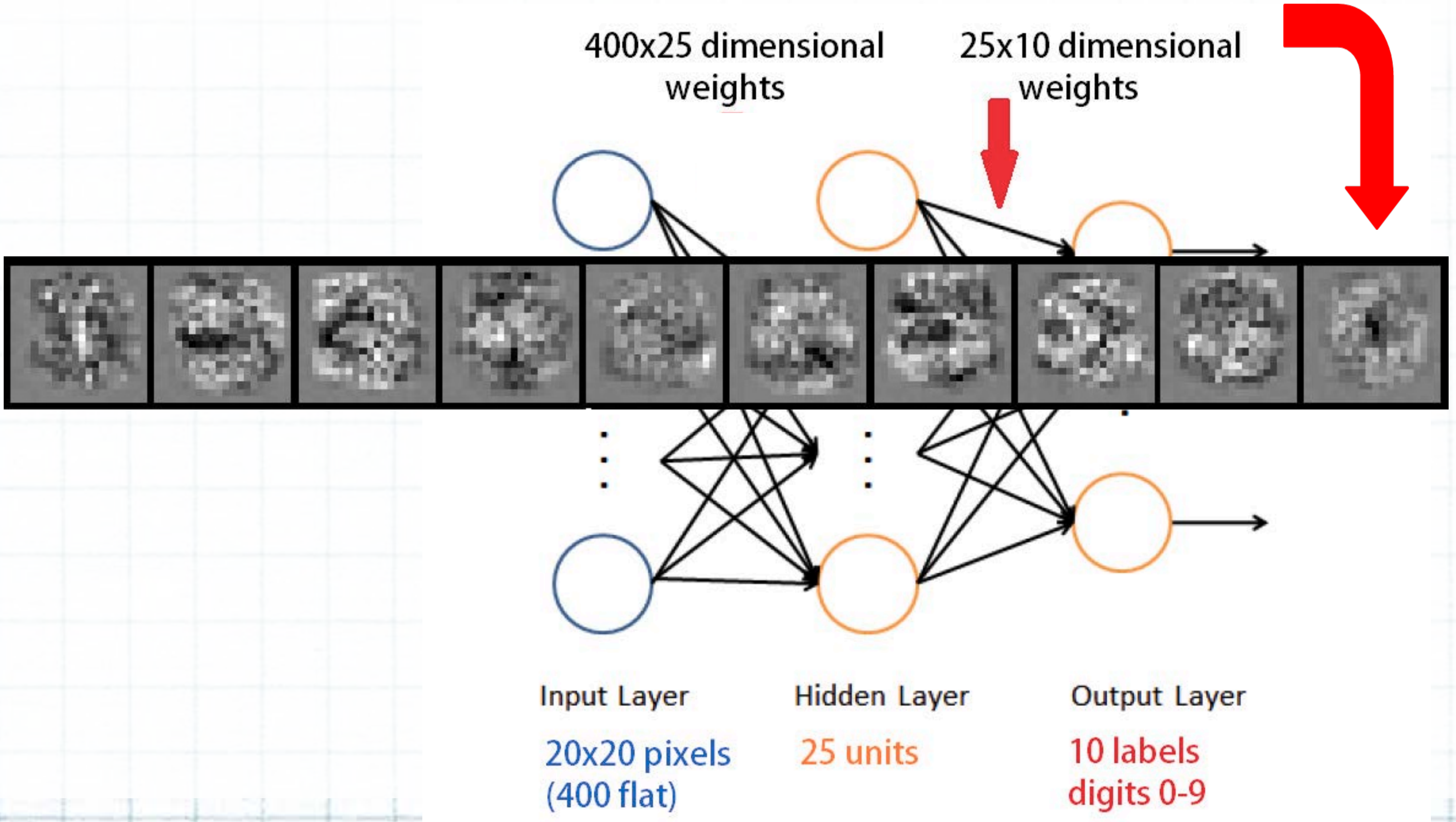




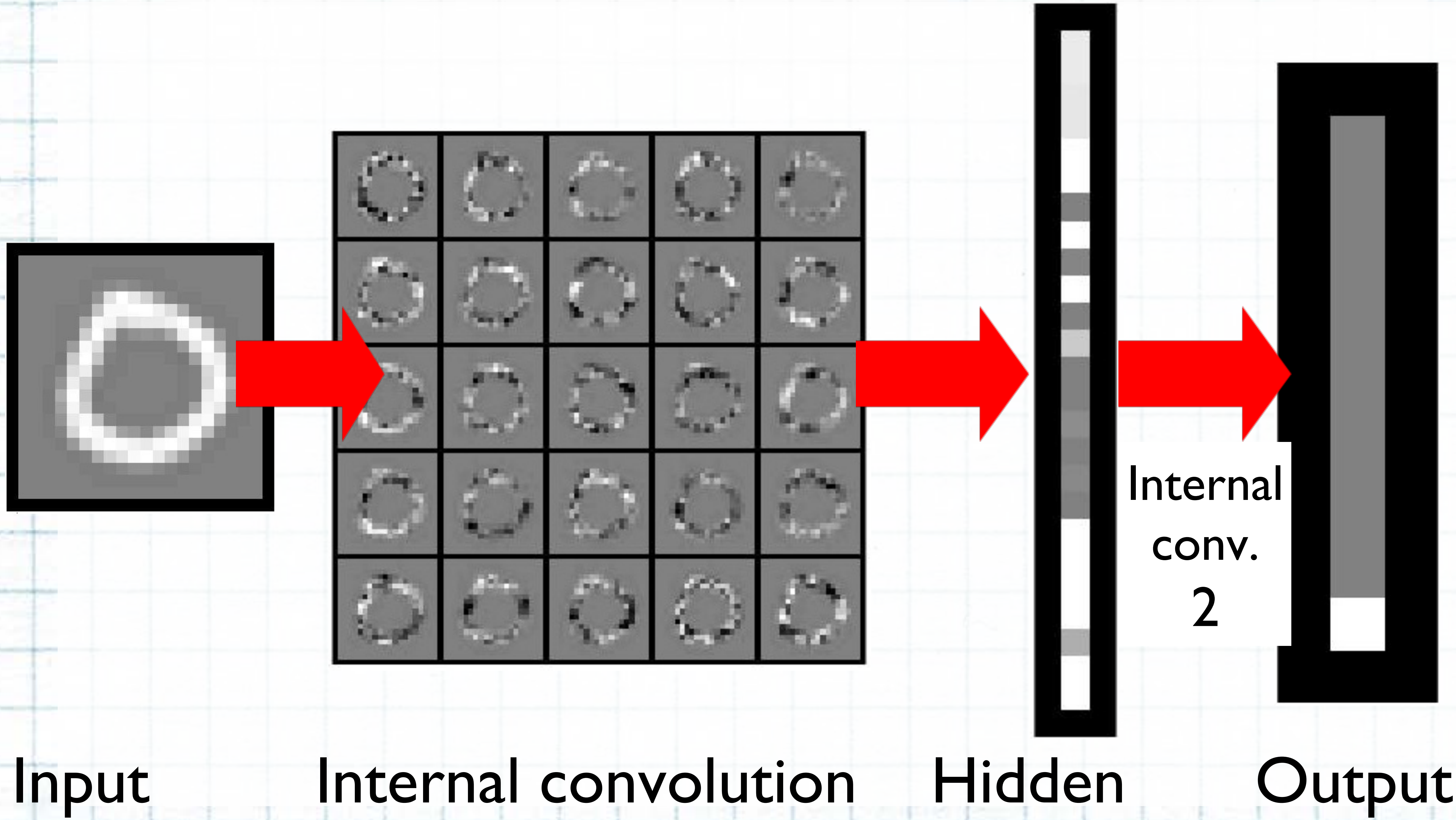


Neural Net Example – Digit Recognition

Hidden units to output

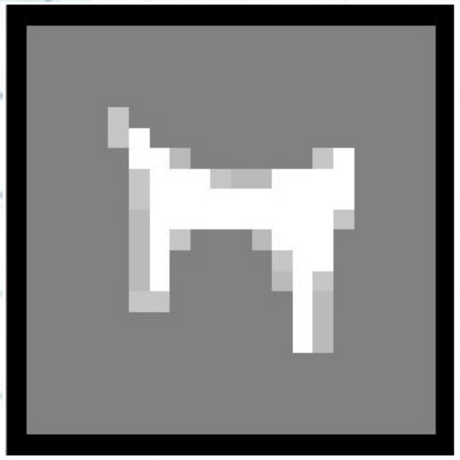


Neural Net Example – Digit Recognition



Neural Net Example – Digit Recognition

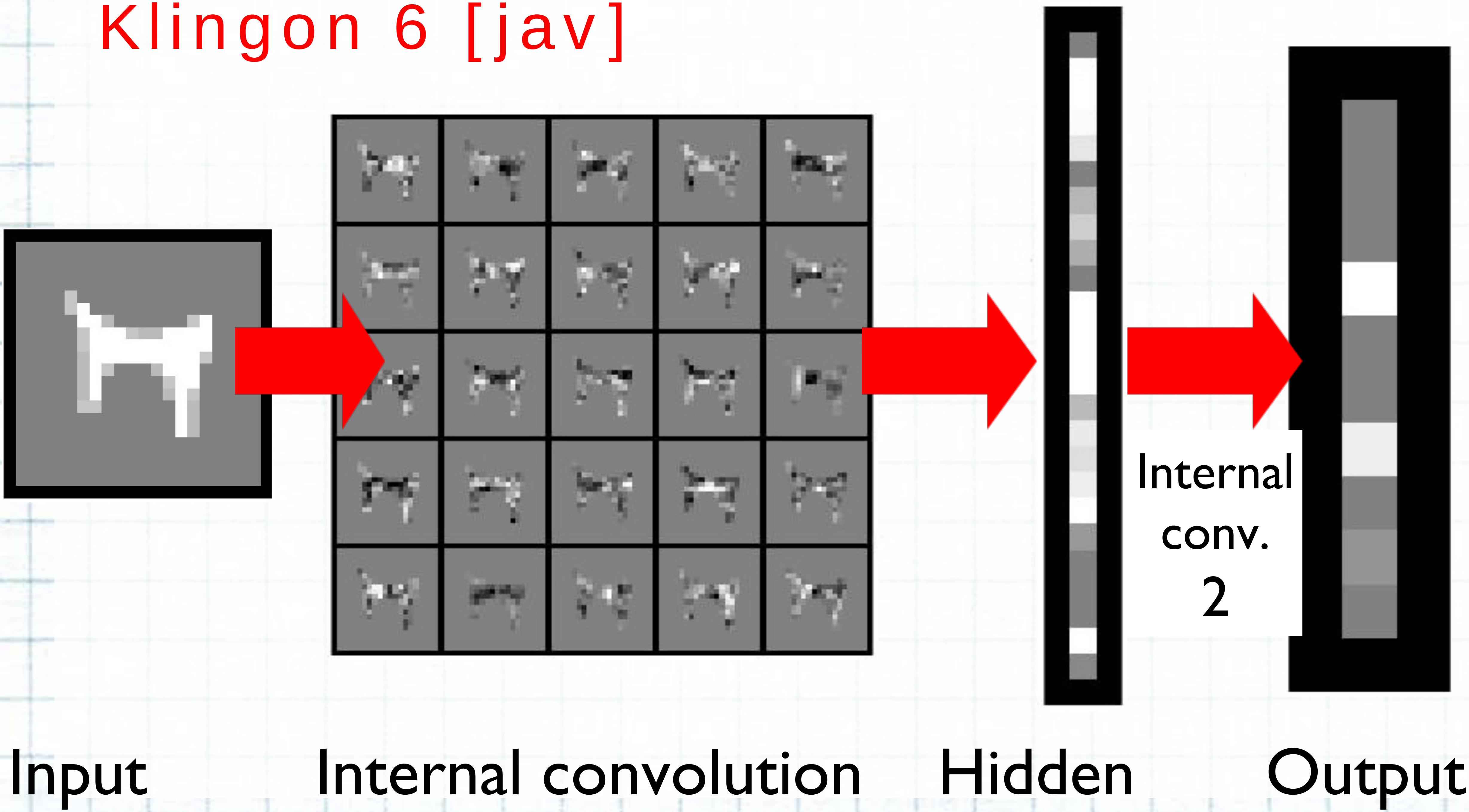
What happens with unknowns?
Klingon 6 [jav]



Input Internal convolution Hidden Output

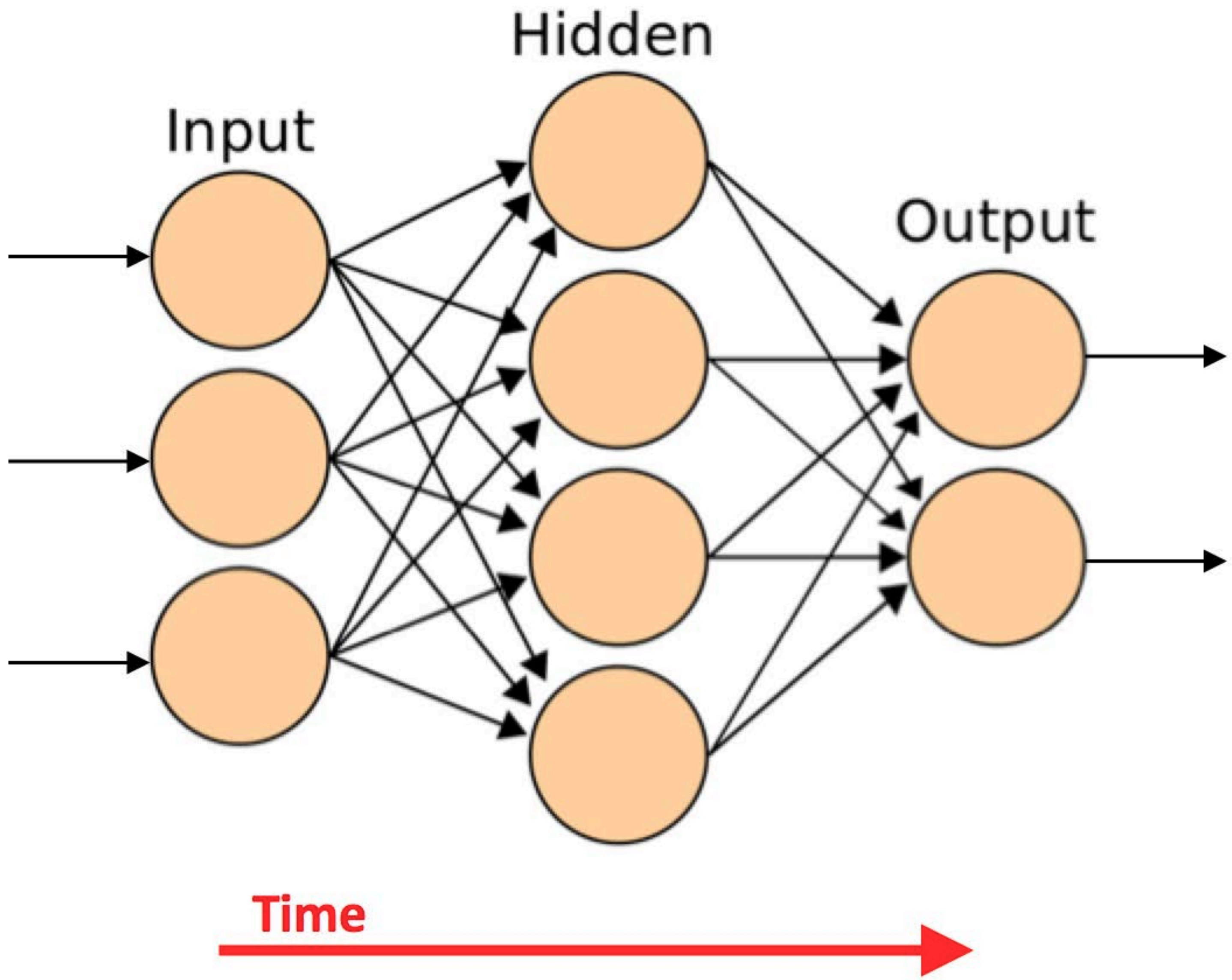
Neural Net Example – Digit Recognition

Klingon 6 [jav]



Artificial Neural Network

Feed-forward neural net



Adapted from: Cburnett, https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

How Do Words Get Into and Out of the Network?

- Challenges for NMT:

→ Input and output length not fixed, different sentence ordering in source and target languages =>
Use recurrent neural networks with attention or convolutional networks

→ Context

→ Training metrics

How Do Words Get Into and Out of the Network?

- Challenges for NMT:

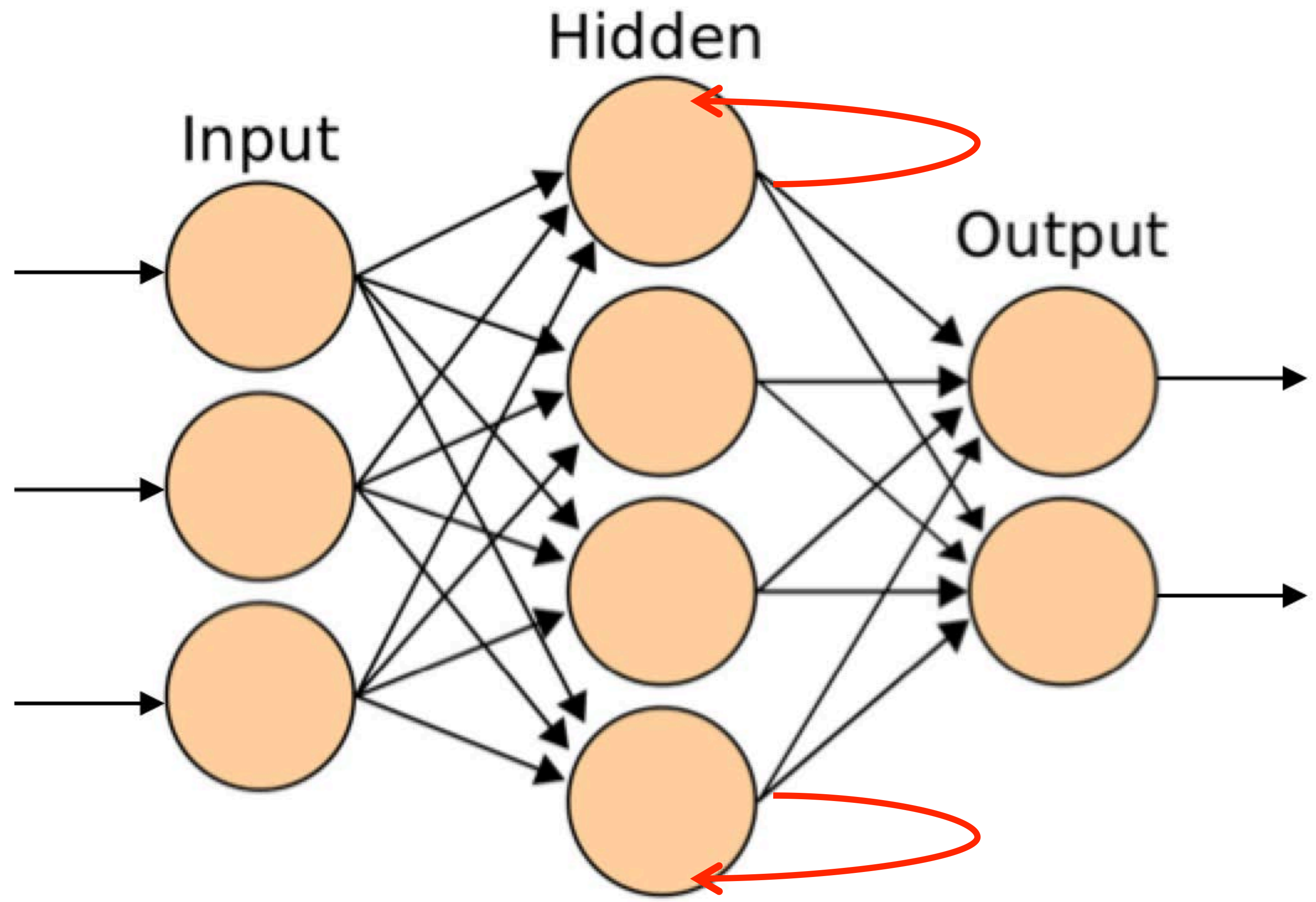
  - Input and output length not fixed, different sentence ordering in source and target languages =>
    Use recurrent neural networks with attention or convolutional networks

  - Context => document (or at least paragraph) level, not sentence level

  - Training metrics

How Do Words Get Into and Out of the Network?

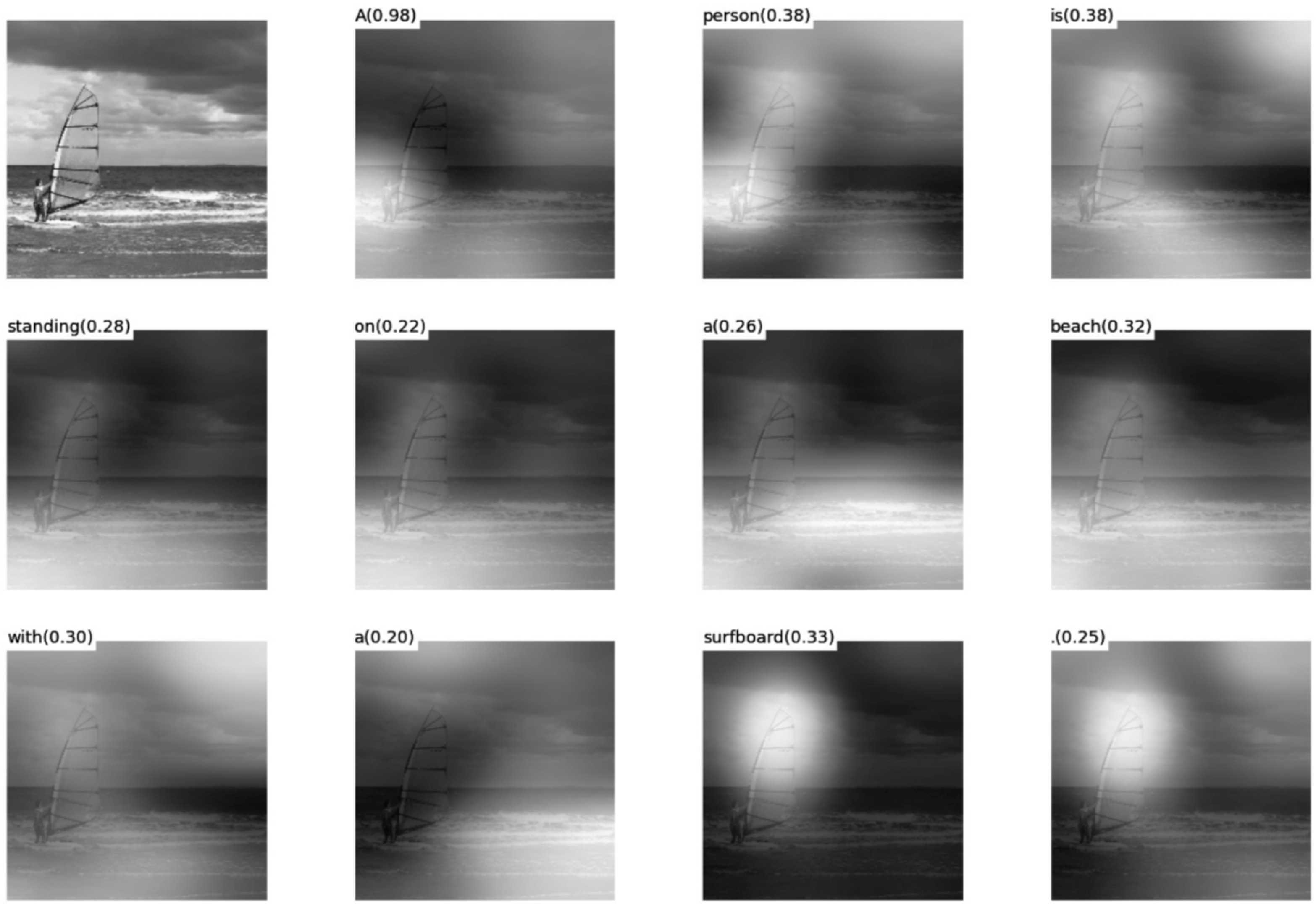
Recurrent neural net



Adapted from: Cburnett, https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

How Do Words Get Into and Out of the Network?

Attention mechanism

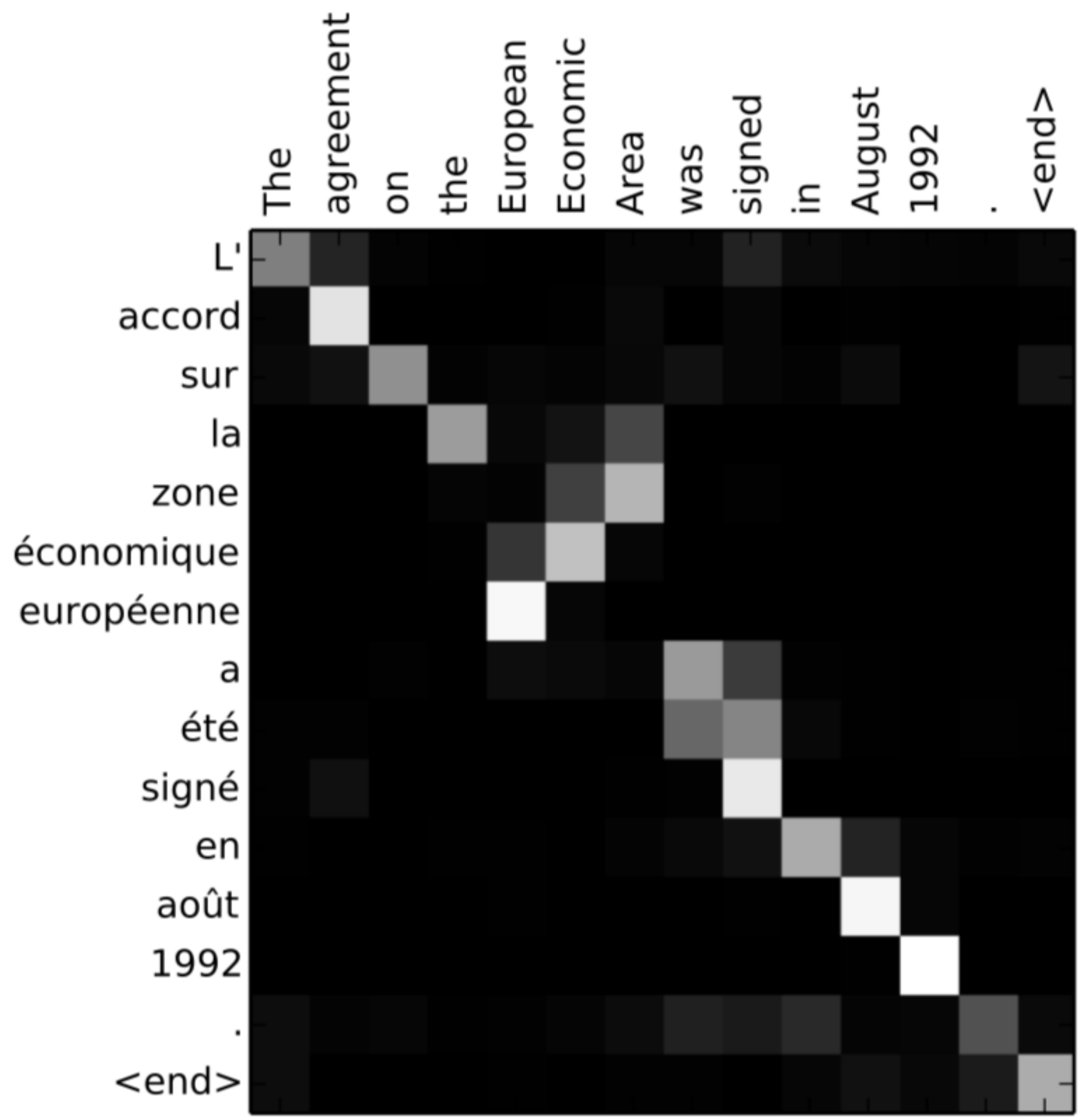


(b) A person is standing on a beach with a surfboard.

Source: K. Xu et al, https://arxiv.org/abs/1502.03044

How Do Words Get Into and Out of the Network?

Attention mechanism



Source: D. Bahdanau et al, ICLR conference proceedings, https://arxiv.org/pdf/1409.0473.pdf

How Do Words Get Into and Out of the Network?

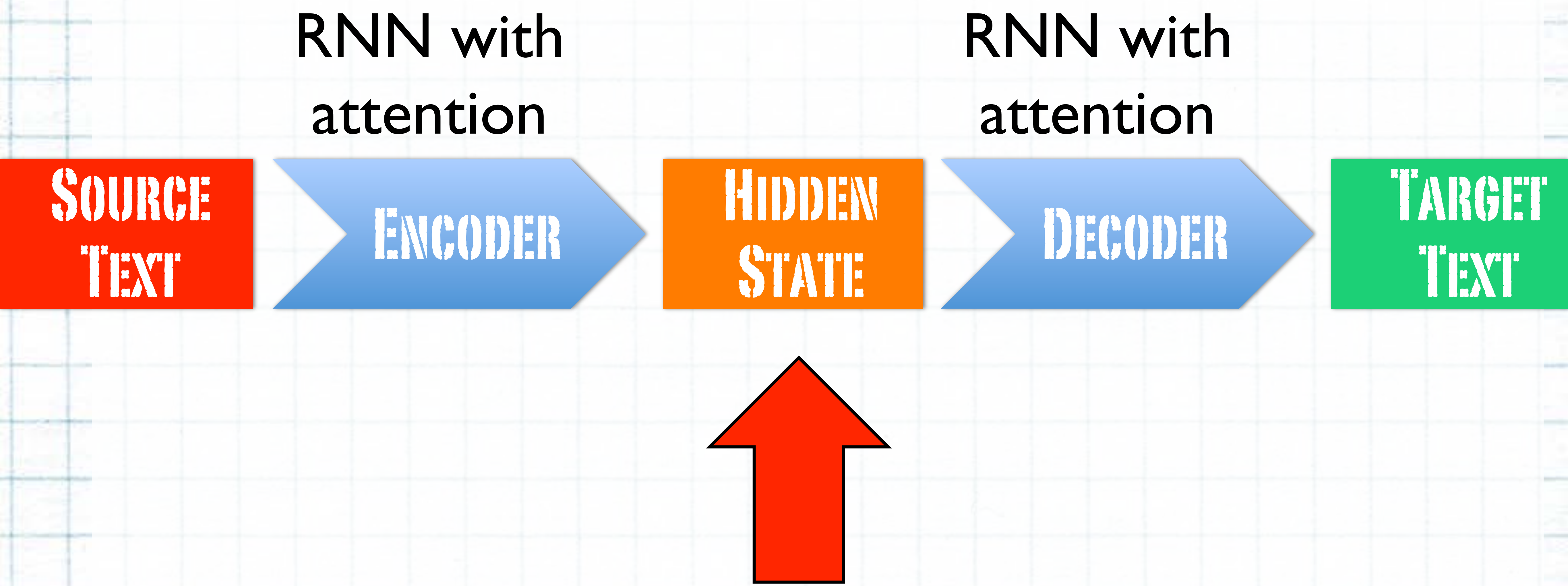
RNN with attention

RNN with attention

Source Text → Encoder → Hidden State → Decoder → Target Text

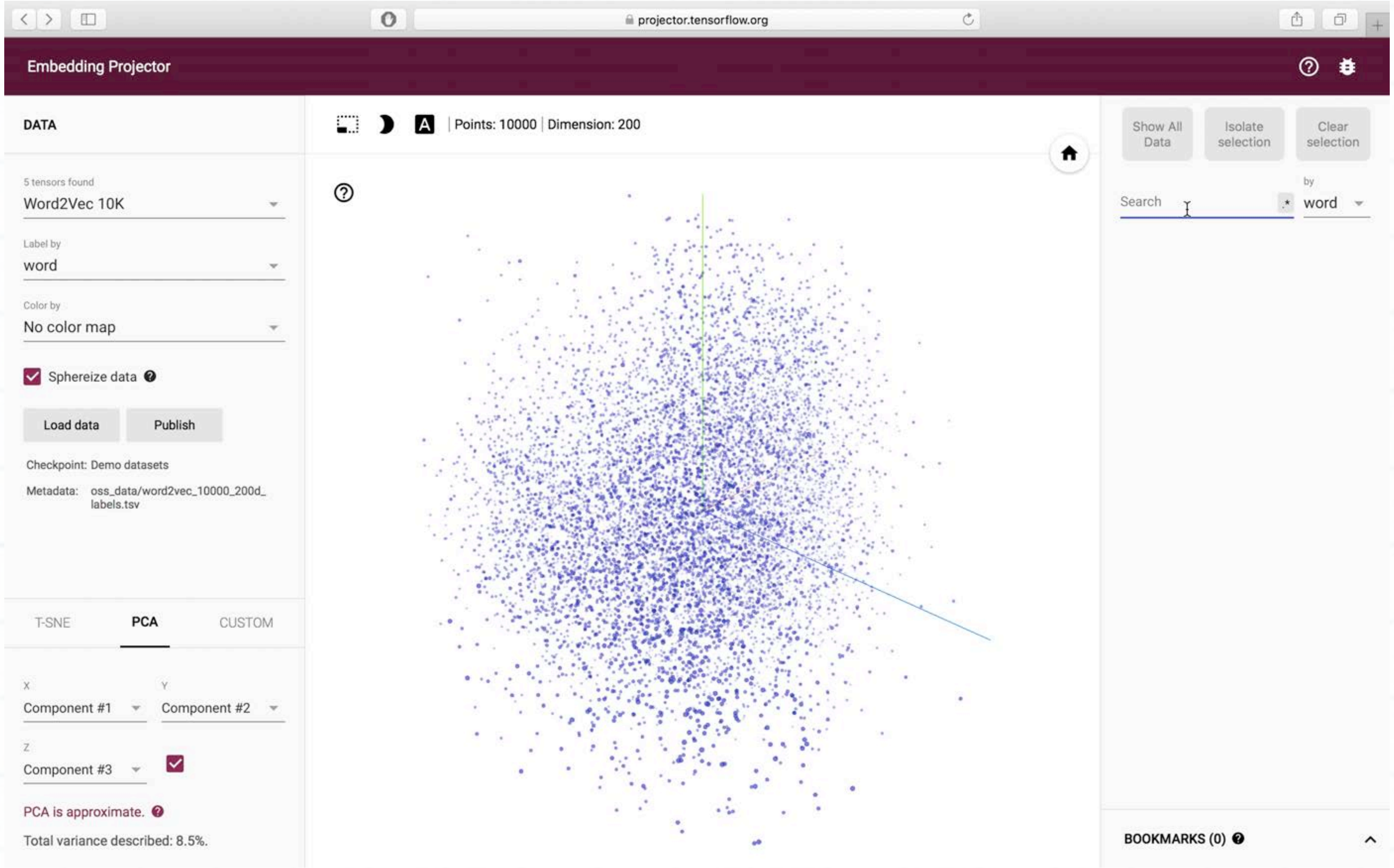
tokenization (optional)

How Do Words Get Into and Out of the Network?



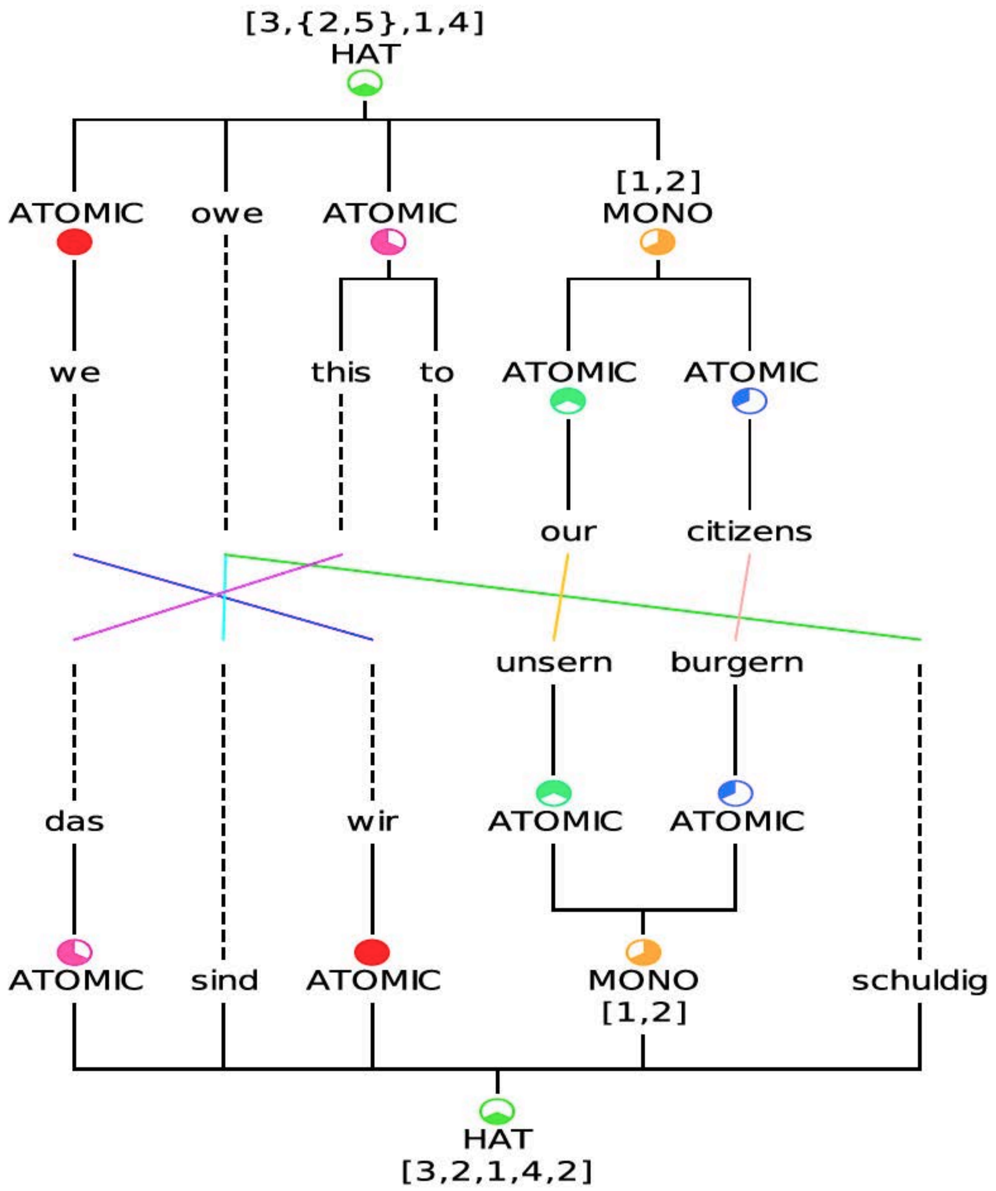
How Do Words Get Into and Out of the Network?

https://projector.tensorflow.org



How Do Words Get Into and Out of the Network?

Recall: SMT



From G. M. de Buy Wenninger, K. Sima'an, PBML No. 101, April 2014, pp. 43

Neural Nets - Recap

- ✓ Training = extraction of "features" (=patterns) from training data

Neural Nets - Recap

- ✓ Training = extraction of "features" (=patterns) from training data
- ✓ The more hidden layers and hidden units, the more parameters (possible overfitting!)

Neural Nets - Recap

- ✓ Training = extraction of "features" (=patterns) from training data
- ✓ The more hidden layers and hidden units, the more parameters (possible overfitting!)
- ✓ Beware: Garbage in -> worse garbage out!

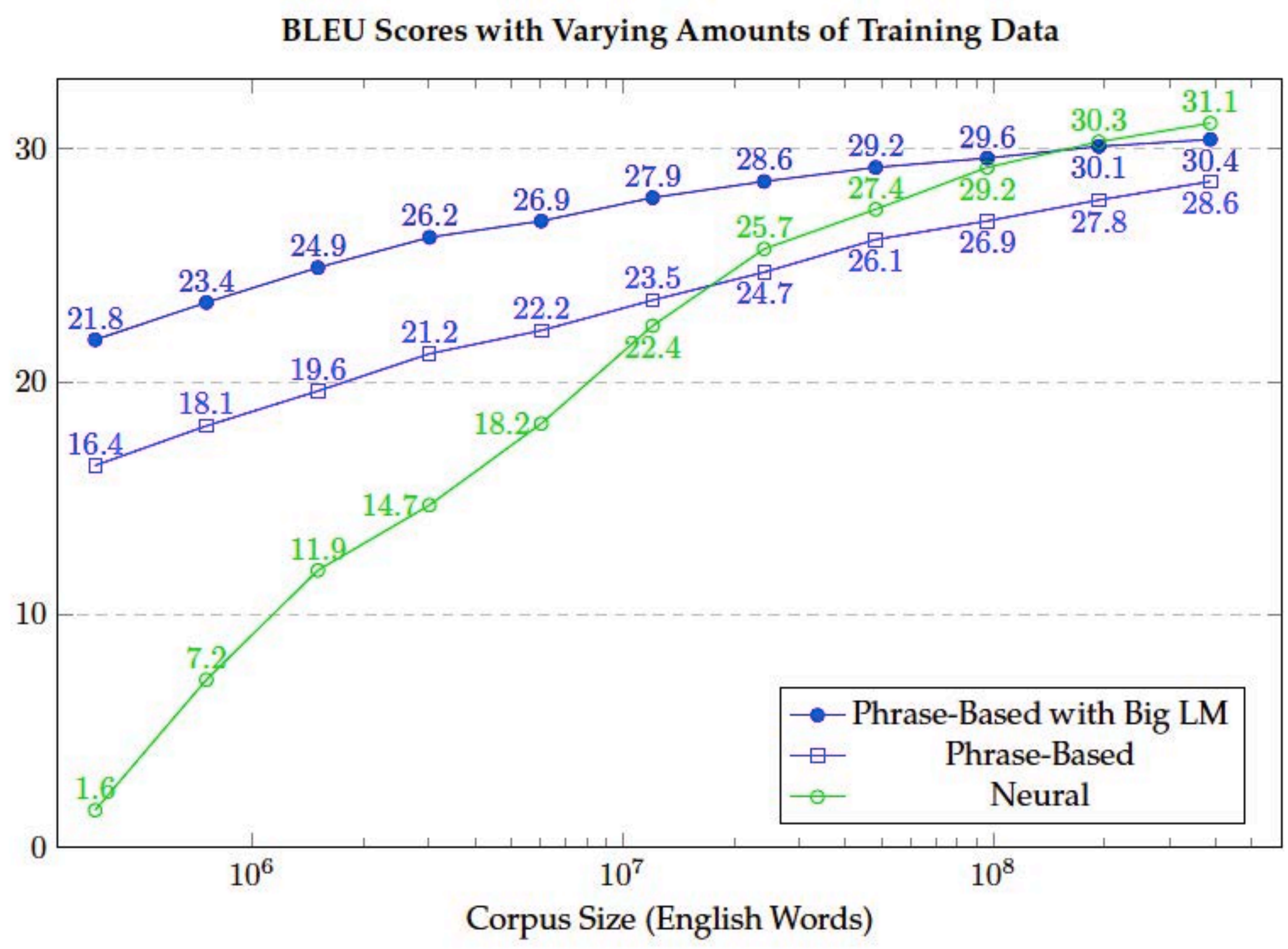
Neural Nets - Recap

- ✓ Training = extraction of "features" (=patterns) from training data
- ✓ The more hidden layers and hidden units, the more parameters (possible overfitting!)
- ✓ Beware: Garbage in -> worse garbage out!
- ✓ ANNs work well for pattern recognition after training, including "context"

Neural Nets - Recap

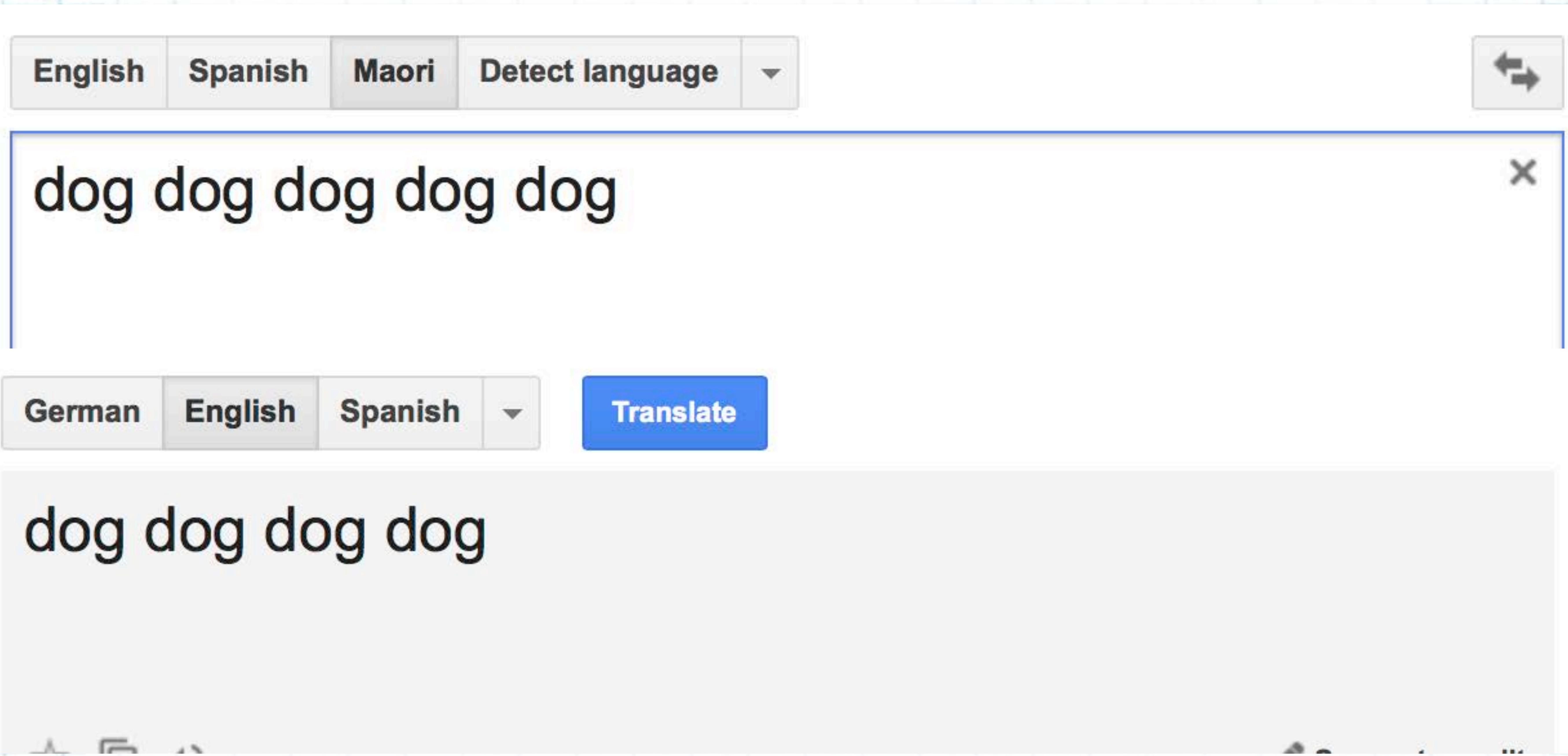
- ✓ Training = extraction of "features" (=patterns) from training data
- ✓ The more hidden layers and hidden units, the more parameters (possible overfitting!)
- ✓ Beware: Garbage in -> worse garbage out!
- ✓ ANNs work well for pattern recognition after training, including "context"
- ✓ Completely unpredictable when confronted with new, hitherto unknown data

Training Data

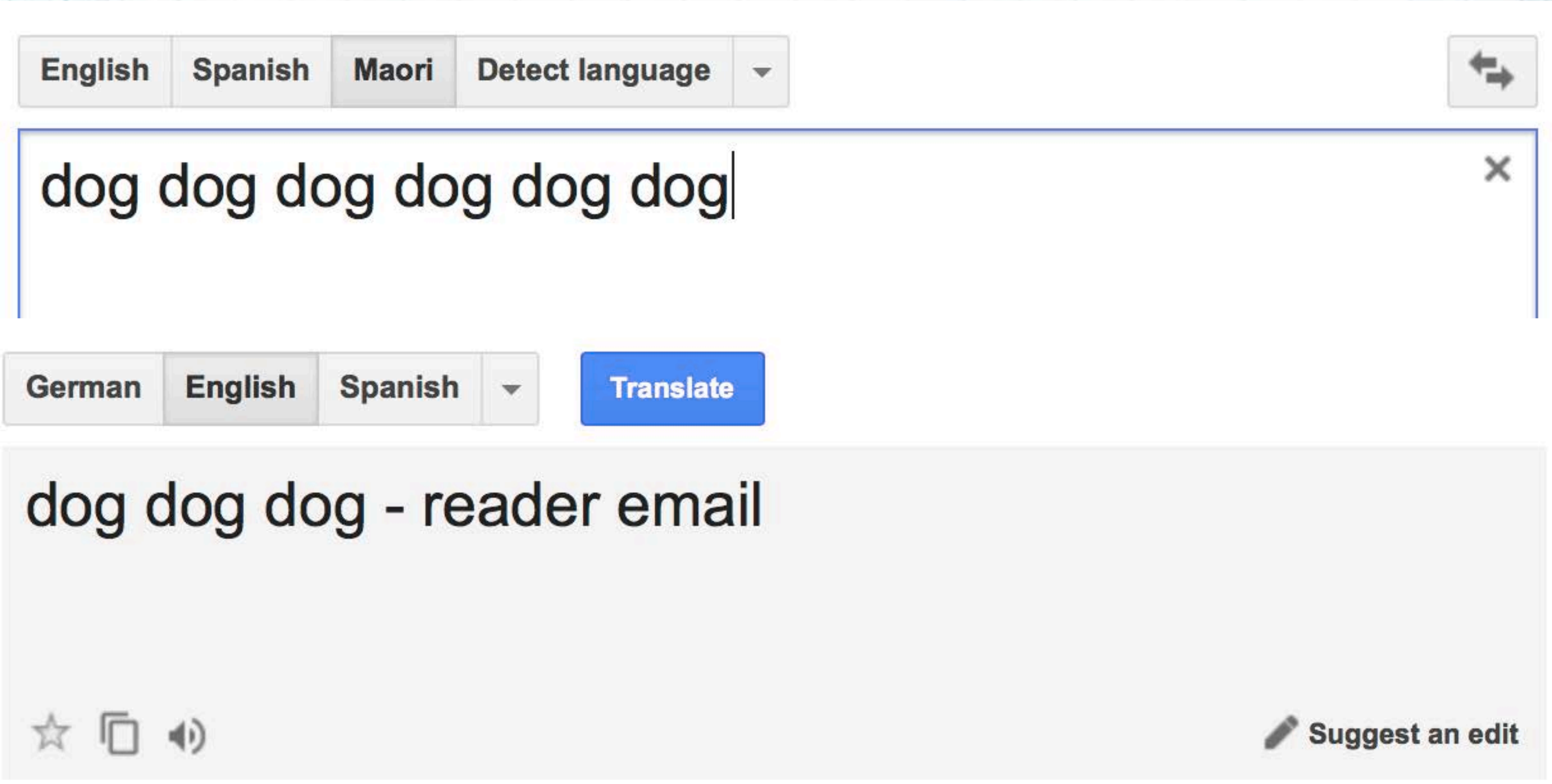


Source: P. Koehn, https://arxiv.org/abs/1709.07809

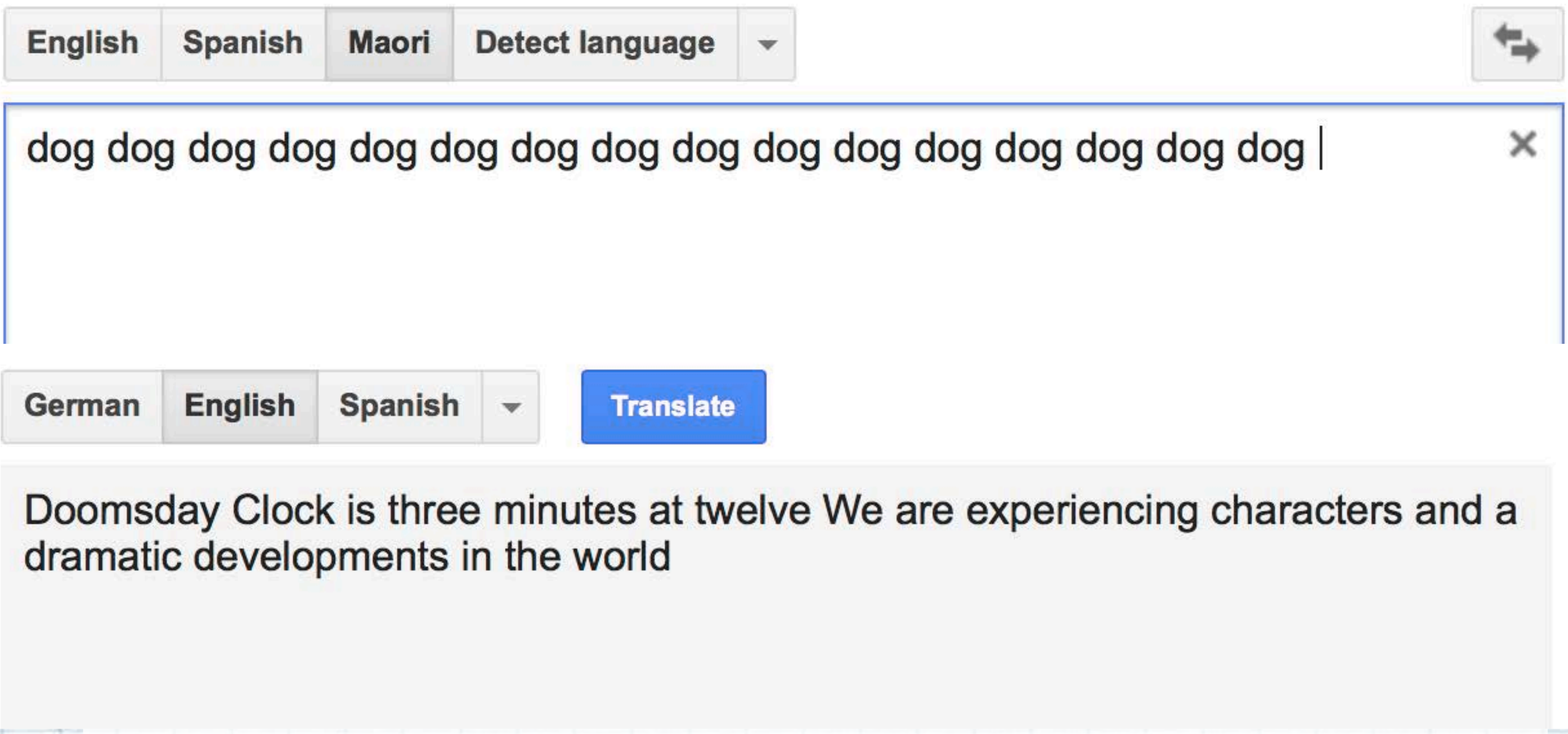
Unpredictability



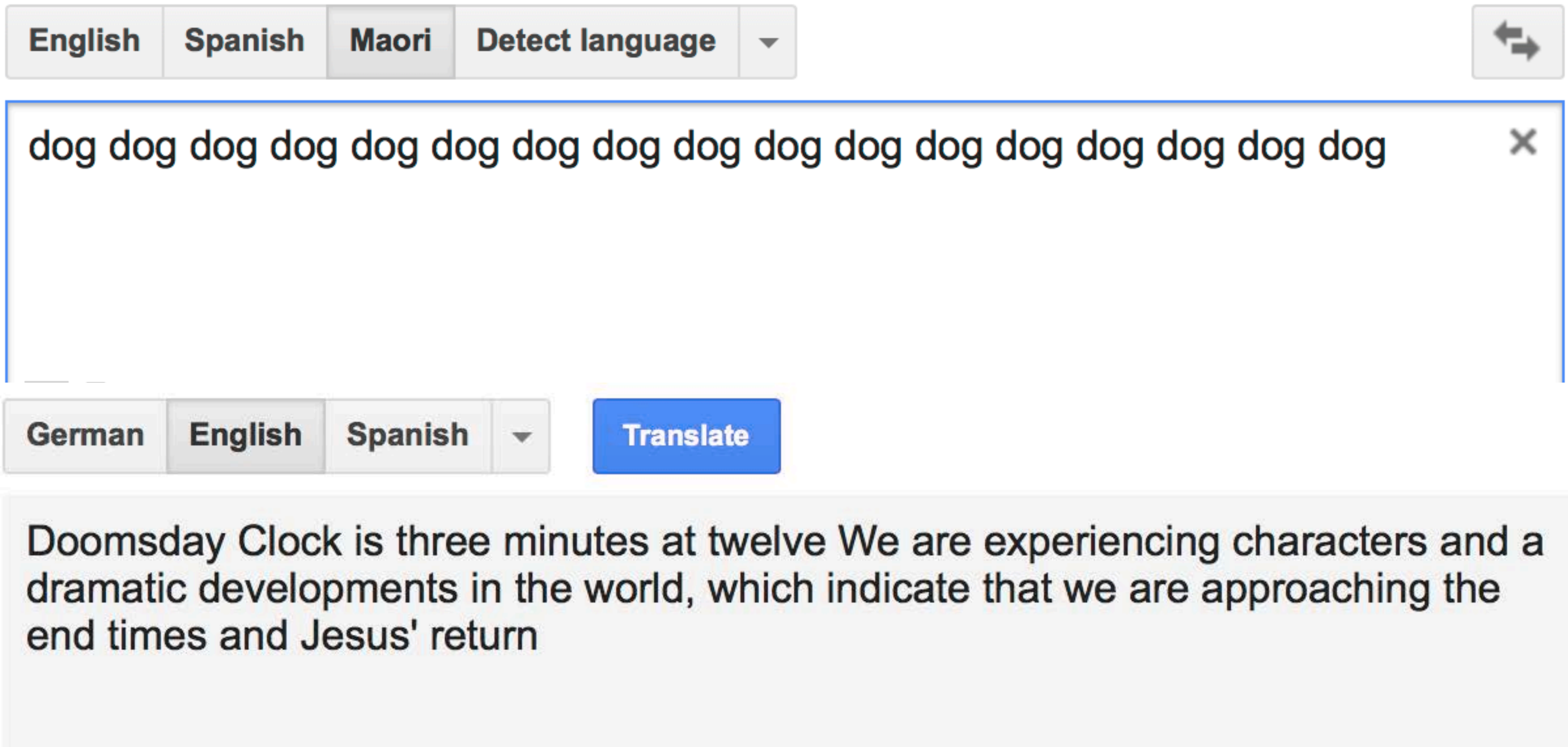
Unpredictability



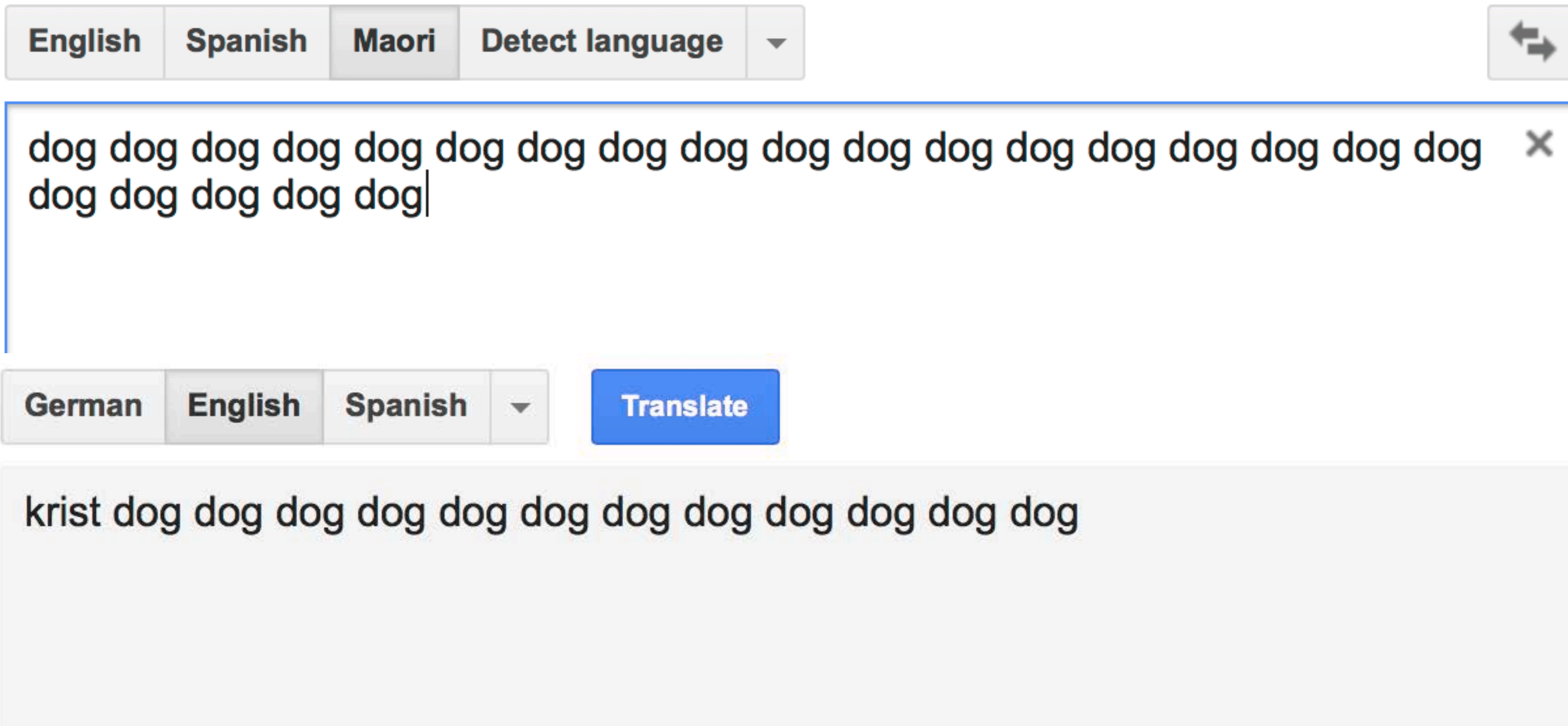
Unpredictability



Unpredictability



Unpredictability



References & Further Reading

- Slides at: https://www.CFBtranslations.com
- Handouts in app and also at https://www.CFBtranslations.com
- A. Ng, *Machine Learning*, Coursera, https://www.coursera.org/learn/machine-learning
- Google's Tensorflow: https://www.tensorflow.org/
- Madly Ambiguous - game to illustrate how NMT deals with context: http://madlyambiguous.osu.edu