ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

ADVANCED DATA SCIENCE TRAINING I

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

OUTLINE

1. A Brief History of A.I.
2. Basics of Neural Networks
3. Case Study: Japanese Pharmacy Chain
4. Deep Learning

LEARNING OBJECTIVES

1. Awareness of some highlights of the recent history of A.I. research
2. Able to define, generally, what an artificial neural network is
3. Understanding of the role of back propagation in training neural nets
4. Awareness of some strengths and limitations of neural networks
5. Able to define, in very general terms, the concept of 'deep learning'

[Yes, Androids Do Dream of Electric Sheep, The Guardian UK]

A BRIEF HISTORY OF A.I.

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

Q: How many legs does a cat have if you call the tail a leg?
A: Four. Calling the tail a leg doesn't make it a leg.
(old riddle, attributed to Abraham Lincoln)

HEADLINES

"AlphaGo vanquishes world's top Go player, marking A.I.'s superiority over human mind" [*South China Morning Post*, May 27, 2017]

"A Japanese A.I. program just wrote a short novel, and it almost won a literary prize" [*Digital Trends*, March 23, 2016]

"Elon Musk: Artificial intelligence may spark World War III" [*CNET*, September 4, 2017]

"A.I. hype has peaked so what's next?" [*TechCrunch*, September 30, 2017]

[D. Hoftsadter, *Gödel, Escher, Bach*, 1979]

WHAT IS ARTIFICIAL INTELLIGENCE (A.I.)?

What are the **essential qualities and skills** of an intelligence?

- provides flexible responses in various scenarios
- takes advantage of lucky circumstances
- makes sense out of contradictory messages
- recognizes the relative importance of a situation's elements
- finds similarities between different situations
- draws distinctions between similar situations
- comes up with new ideas from scratch or by re-arranging previous known concepts

DISCUSSION

Do you think that Hofstadter's definition is still valid?

WHAT IS ARTIFICIAL INTELLIGENCE?

A.I. research is defined as the study of **intelligent agents**: any device that perceives its environment and takes actions that maximize its chance of success at some goal.

Examples

- Expert Systems

  TurboTax, WebMD, technical support, insurance claim processing, air traffic control, etc.
- Decision-Making

  Deep Blue, auto-pilot systems, "smart" meters, etc.
- Natural Language Proc.

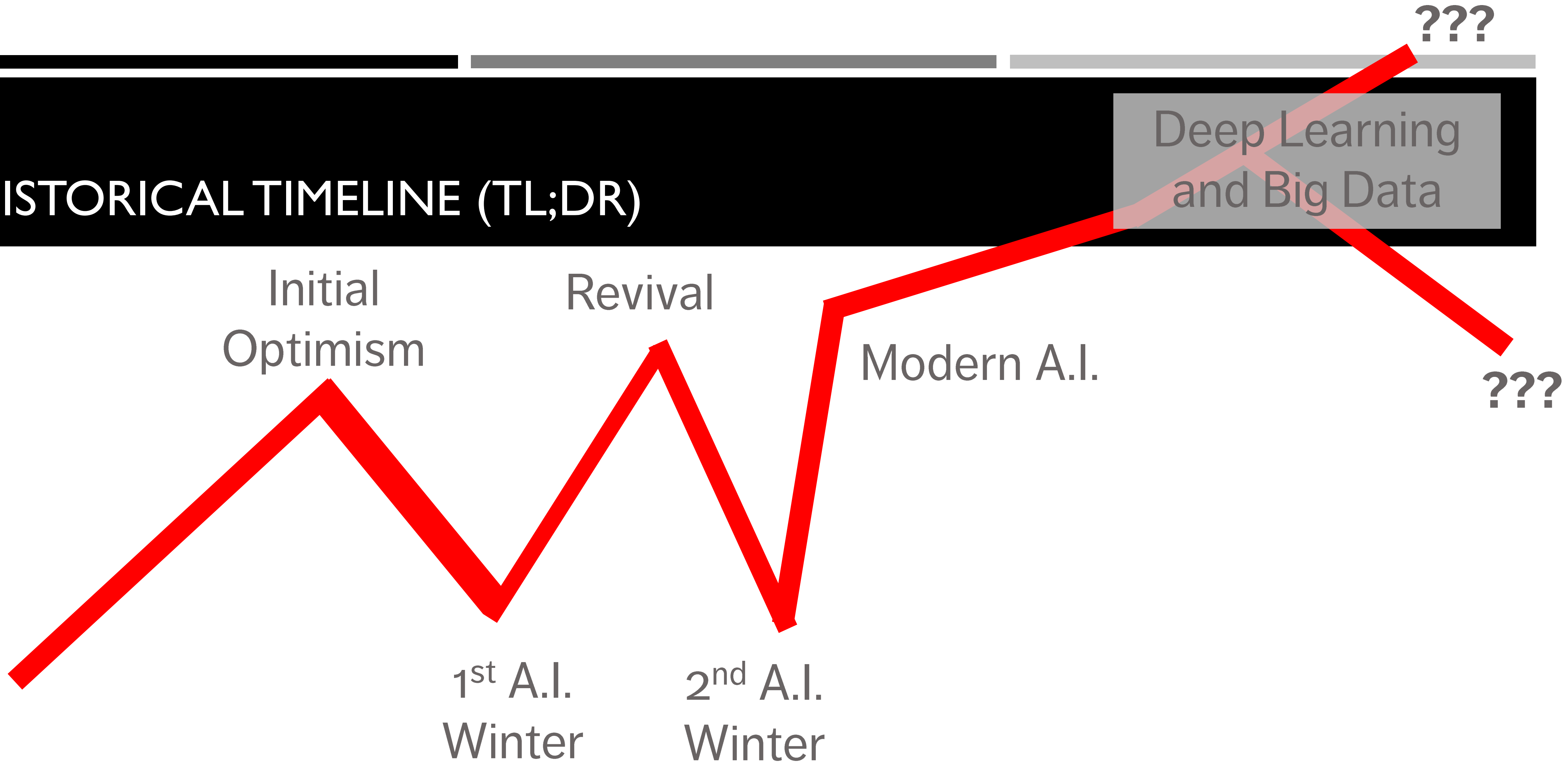
  machine translation, Siri, named-entity recognition, etc.
- Recommenders

  Google, Expedia, Facebook, LinkedIn, Netflix, Amazon, etc.
- Content generators

  music composer, novel writer, animation creator, etc.
- Classifiers

  facial recognition, object identification, fraud detection, etc.

HISTORICAL TIMELINE (TL;DR)



1950s 1960s 1970s 1980s 1990s 2000s 2010s 2020s

Turing Test
Dartmouth Summer
5th Generation Project
Backpropagation
Deep Blue
Human Brain Project
Stanley
???

DISCUSSION

Where do you think A.I. is heading? 3rd A.I. Winter or onto bigger and better things?

How can businesses and governments prepare themselves for either eventuality?

NEURAL NETWORK BASICS

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

"Neural networks blow all previous techniques out of the water in terms of performance, but given the existence of adversarial examples, it shows we really don't understand what's going on."

(D. Gershgorn, *Quartz*)

NEURAL NETWORKS IN A NUTSHELL

A trained **Artificial Neural Network** (ANN) is a function that maps inputs to outputs in a useful way:

- receives input(s)
- computes values
- provides output(s)

ANNs use a Swiss-army-knife approach to things (**plenty of options, but it's not always clear which one should be used**).

The user does not need to decide much about the function or know much about the problem space in advance (**quiet model**).

NEURAL NETWORKS IN A NUTSHELL

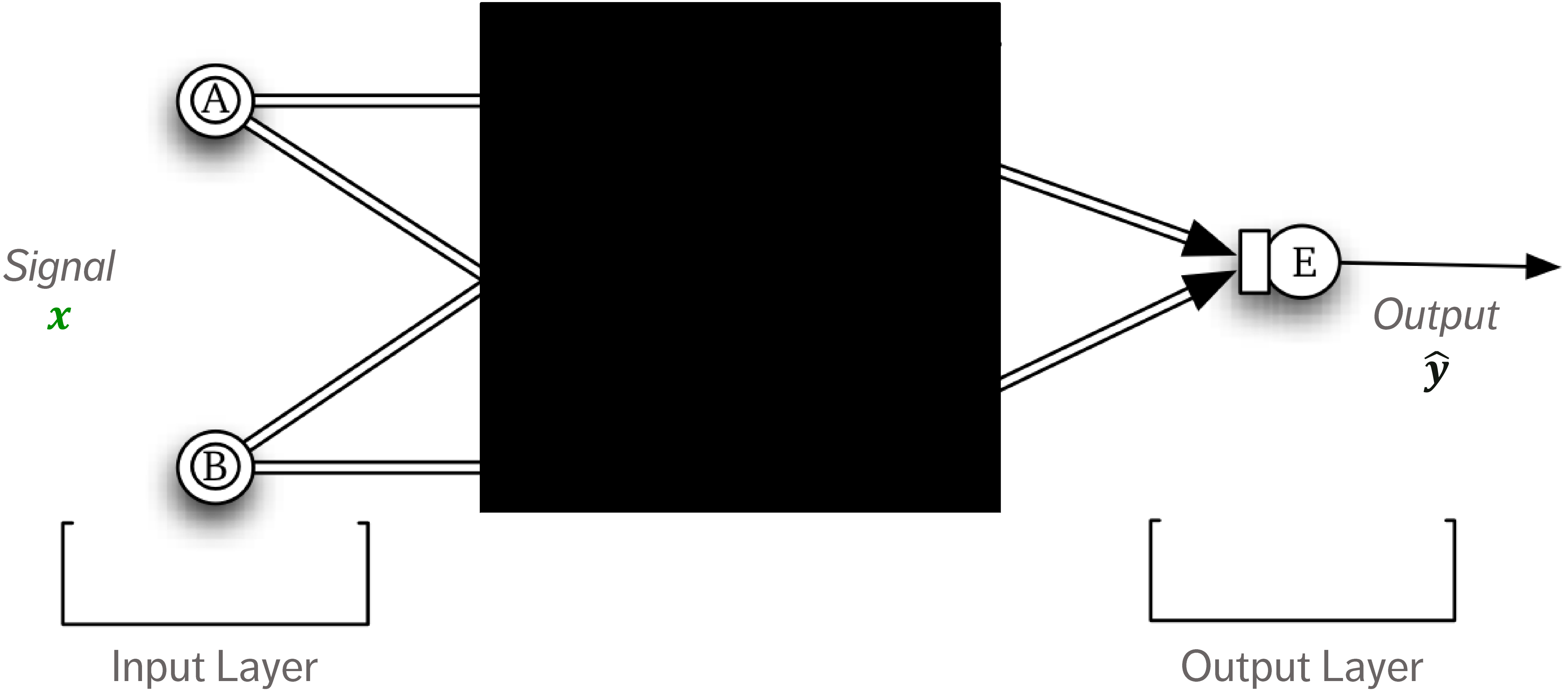
Algorithms allow ANNs to **learn** (i.e. generate the function and its internal values) **automatically**.

ANNs can be used for:

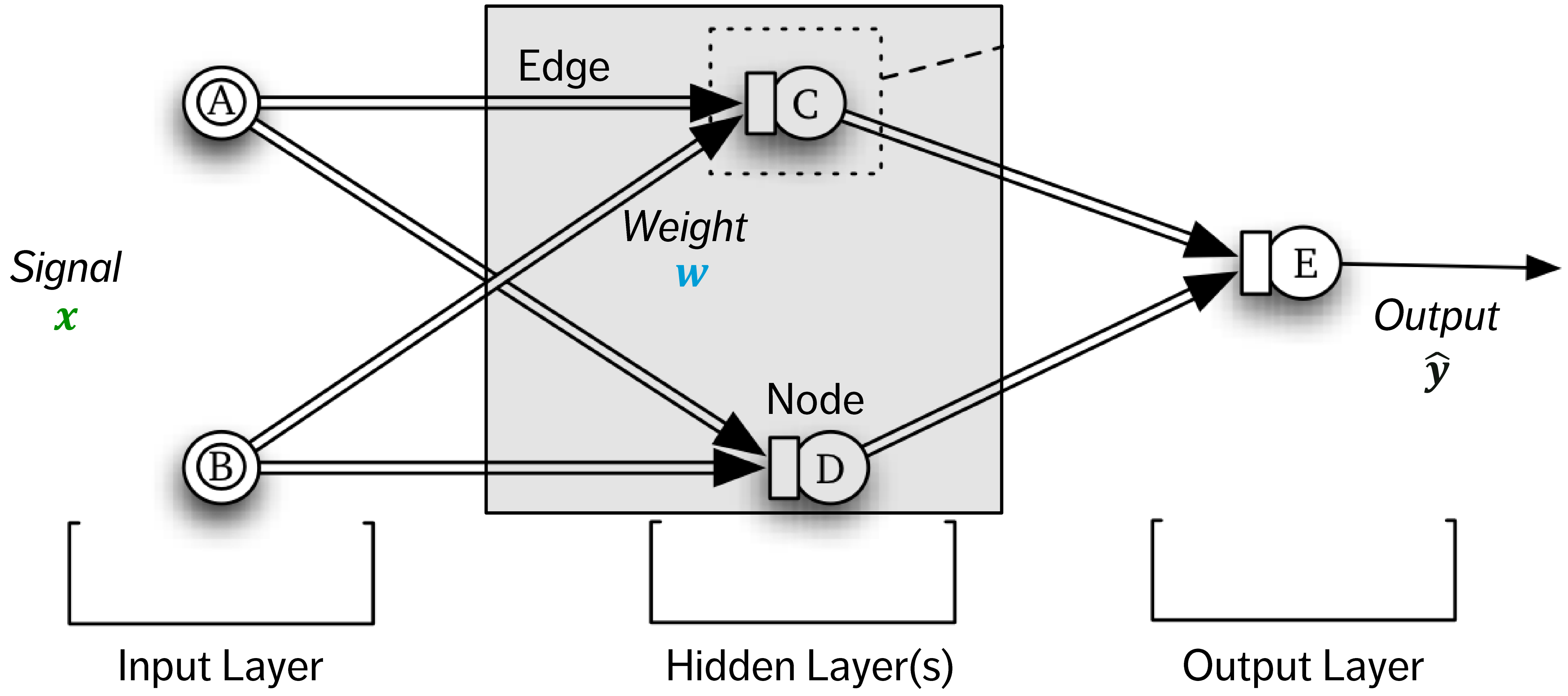
- supervised learning (**multi-layered feedforward neural networks**)
- unsupervised learning (**self-organizing maps**)
- reinforcement learning.

Technically, the only requirement is the ability to minimize a cost function (**optimization**).

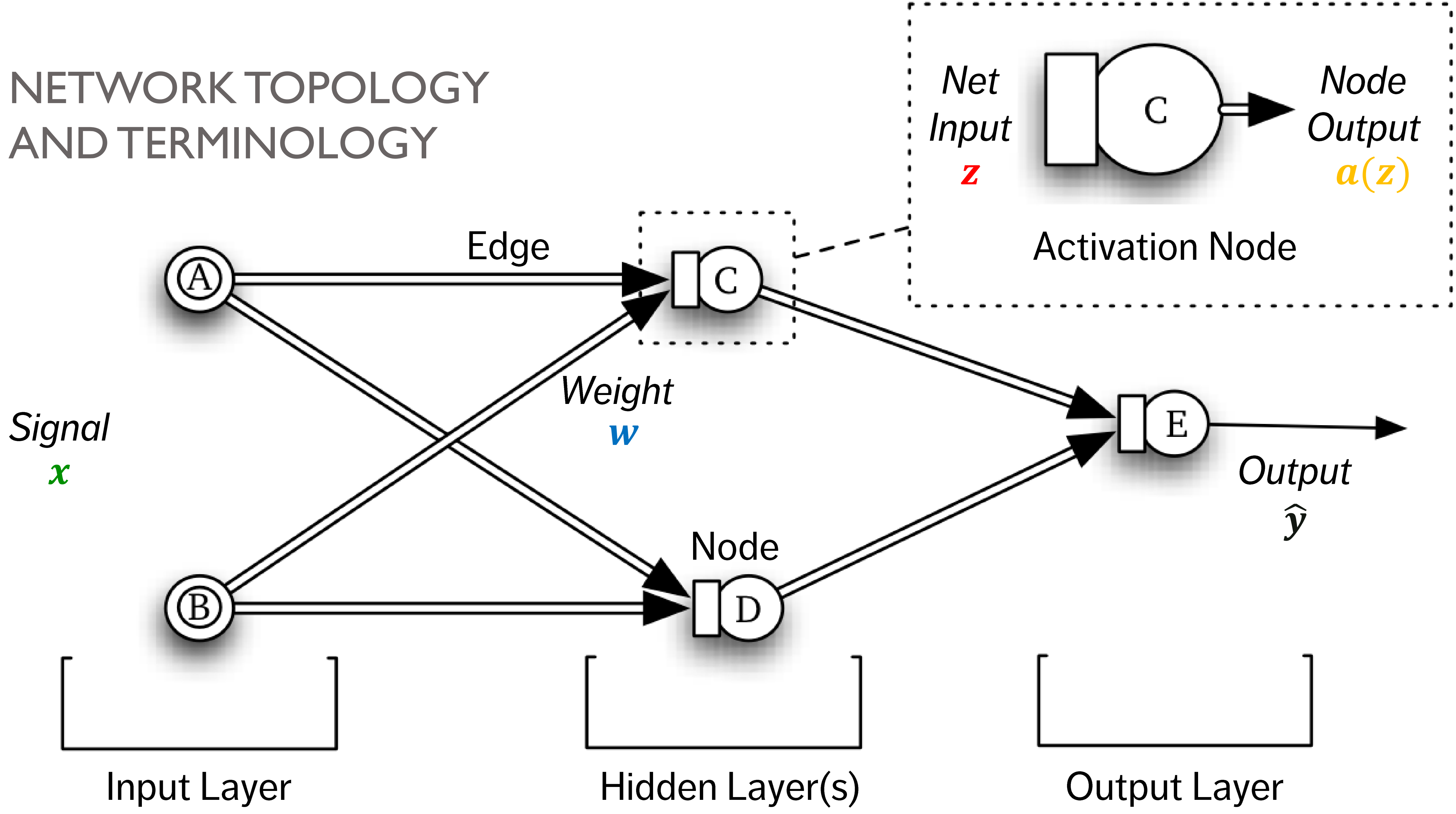
NETWORK TOPOLOGY AND TERMINOLOGY



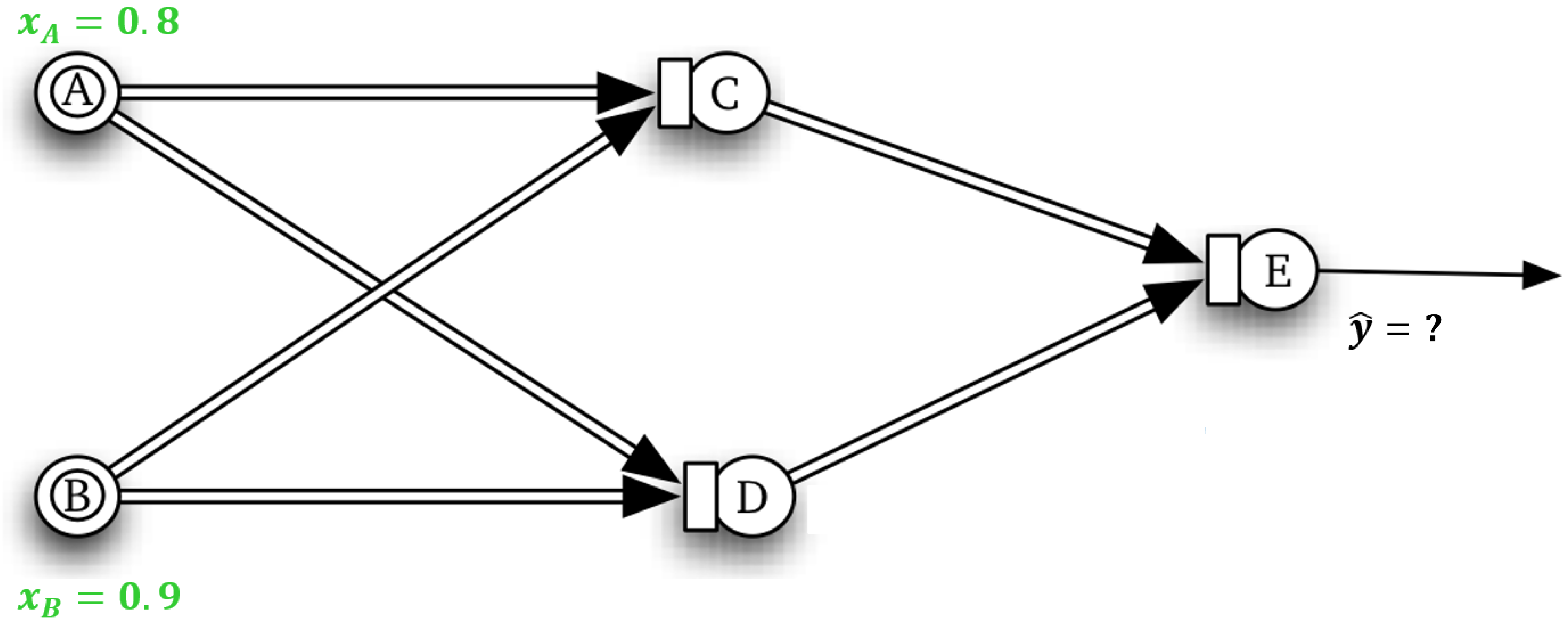
NETWORK TOPOLOGY AND TERMINOLOGY



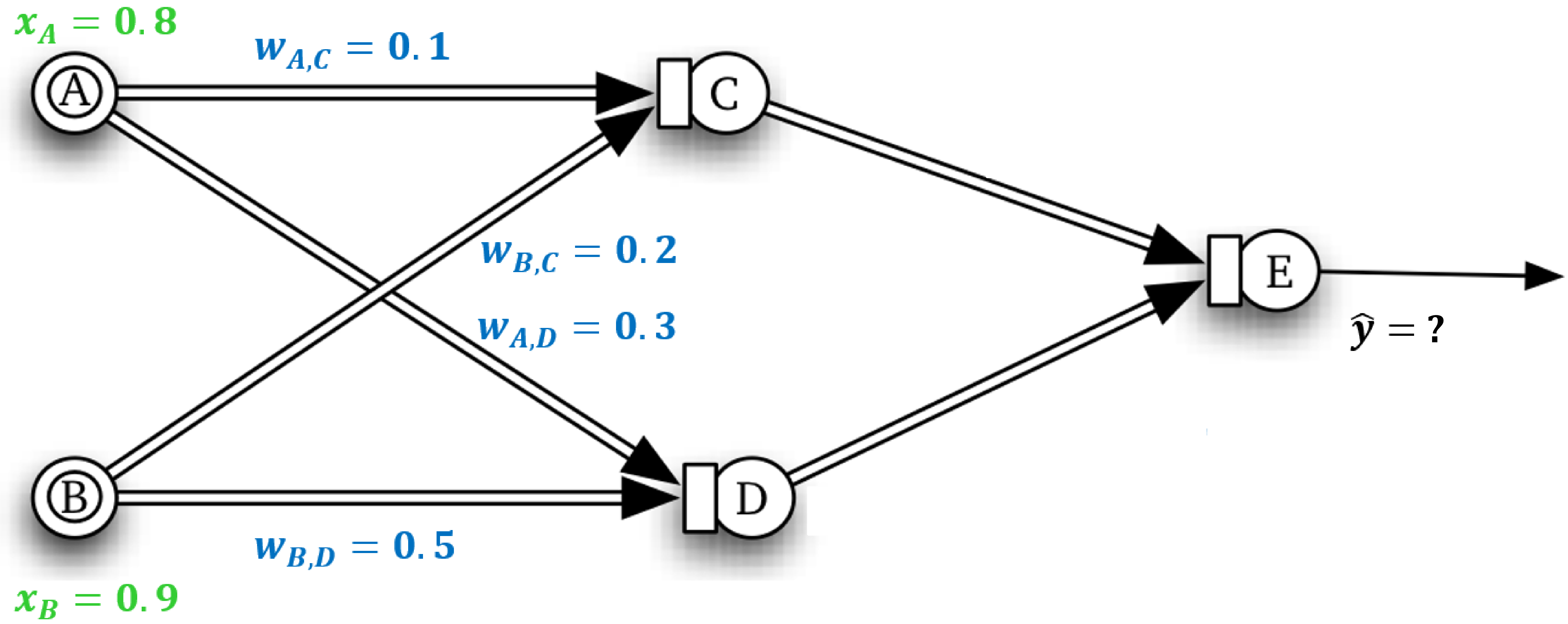
NETWORK TOPOLOGY AND TERMINOLOGY



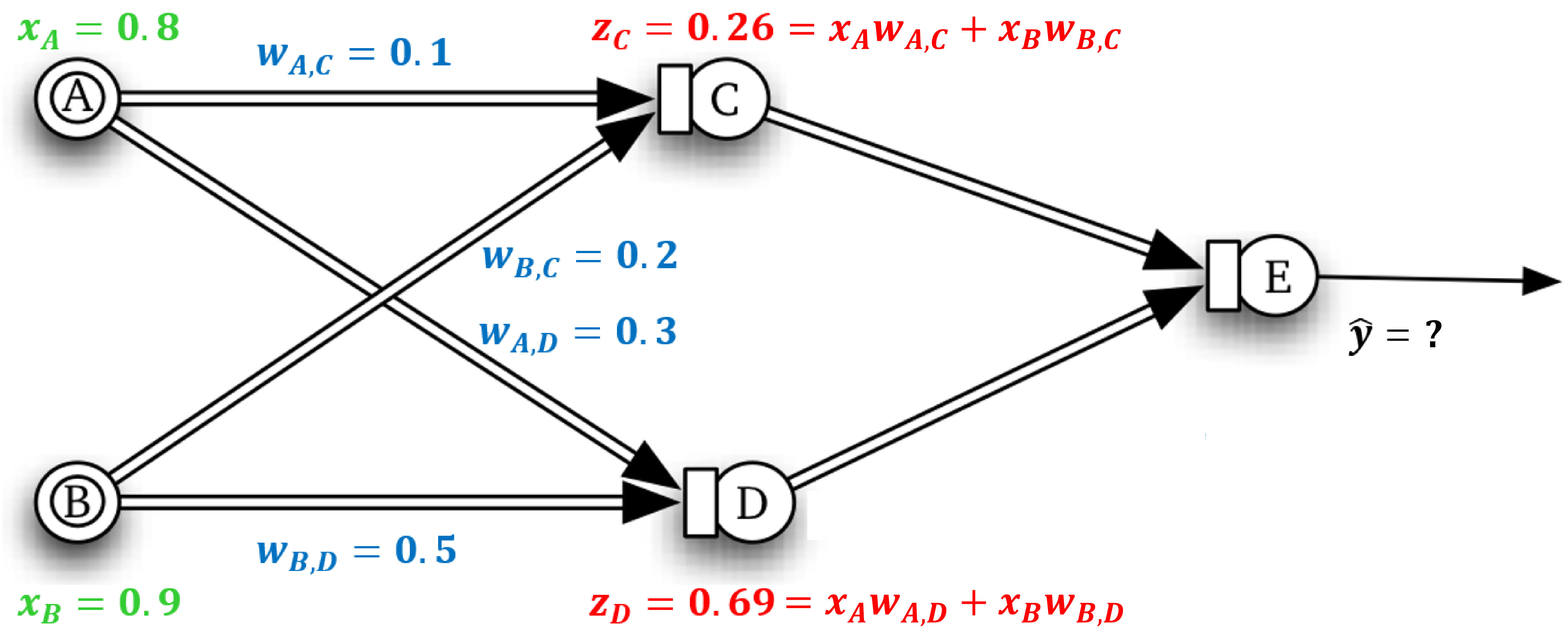
FEED FORWARD NETWORK



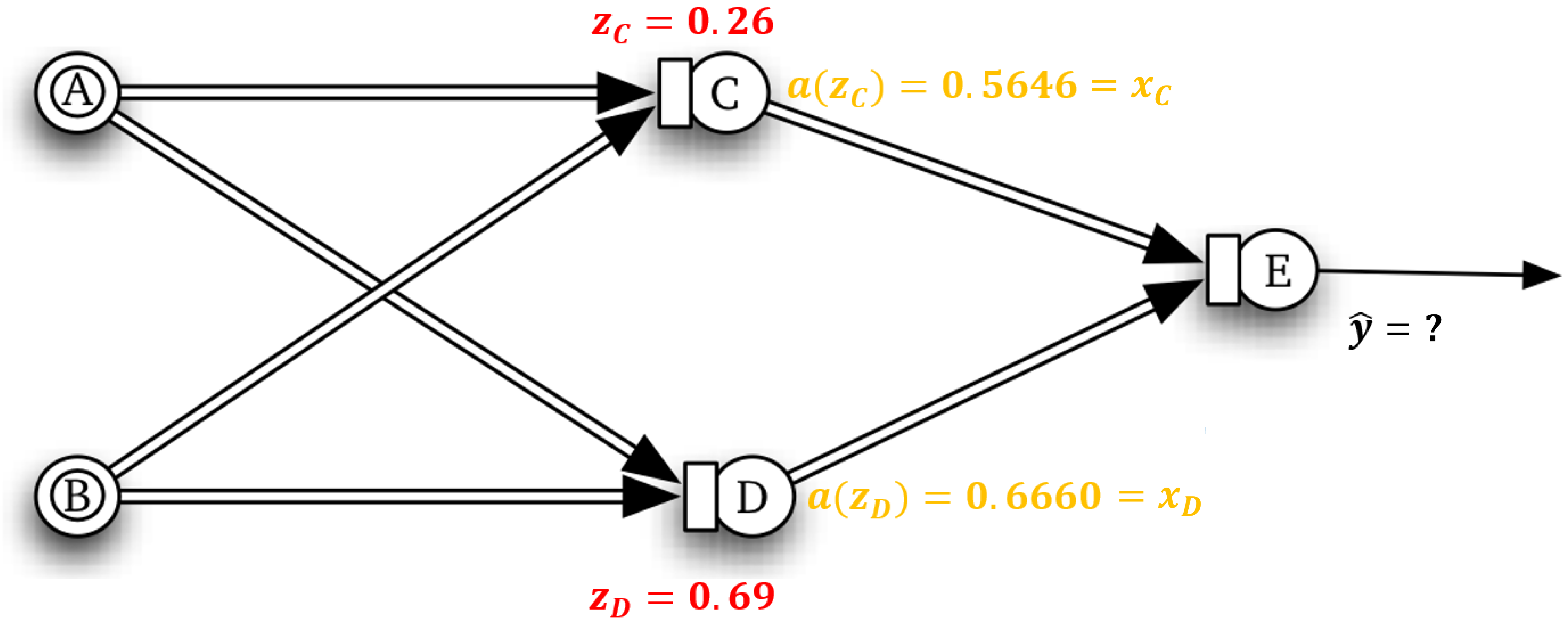
FEED FORWARD NETWORK



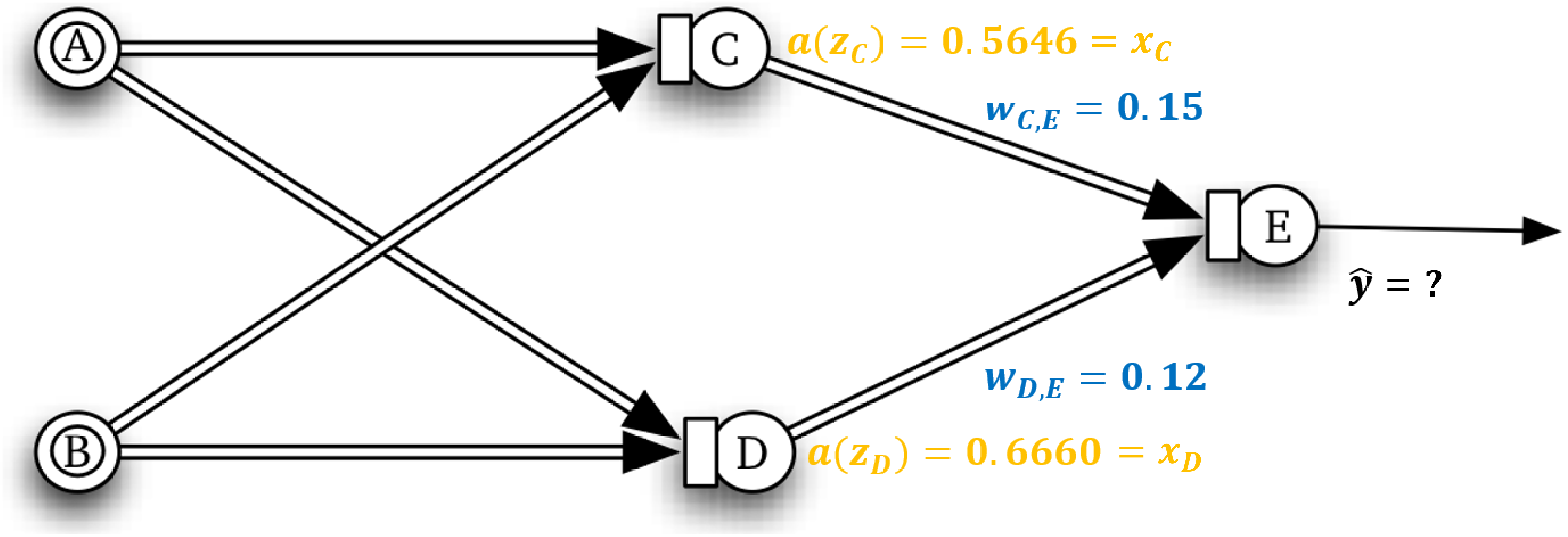
FEED FORWARD NETWORK



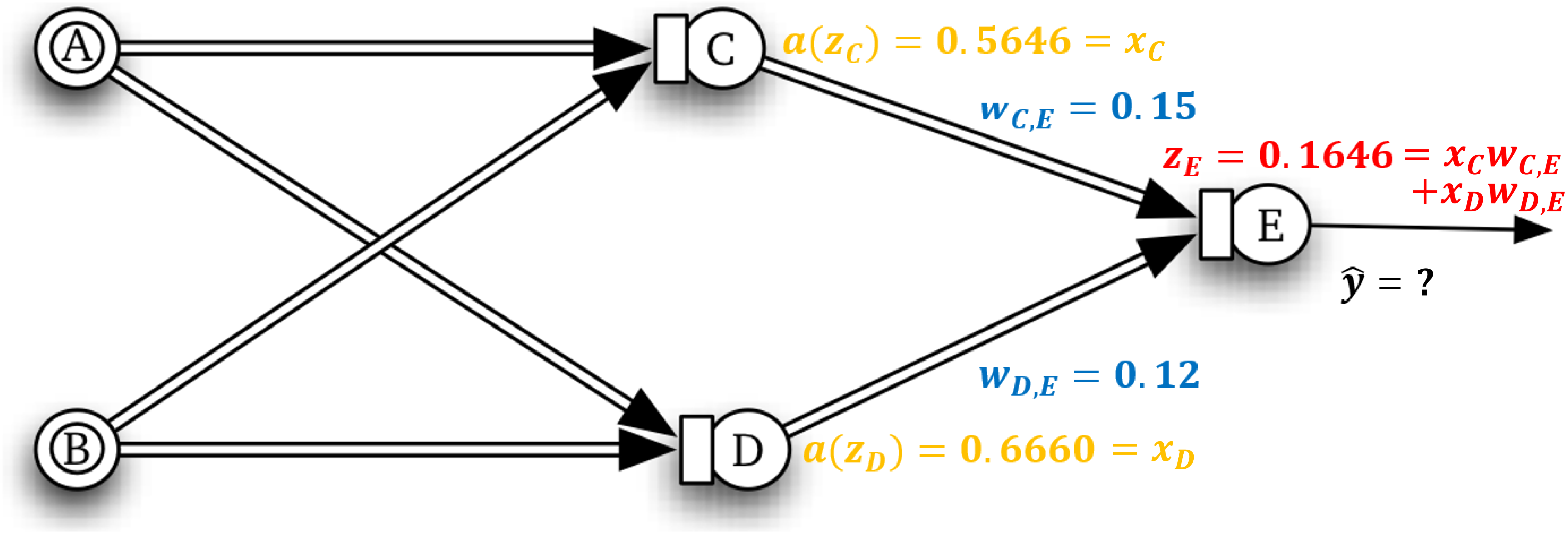
FEED FORWARD NETWORK



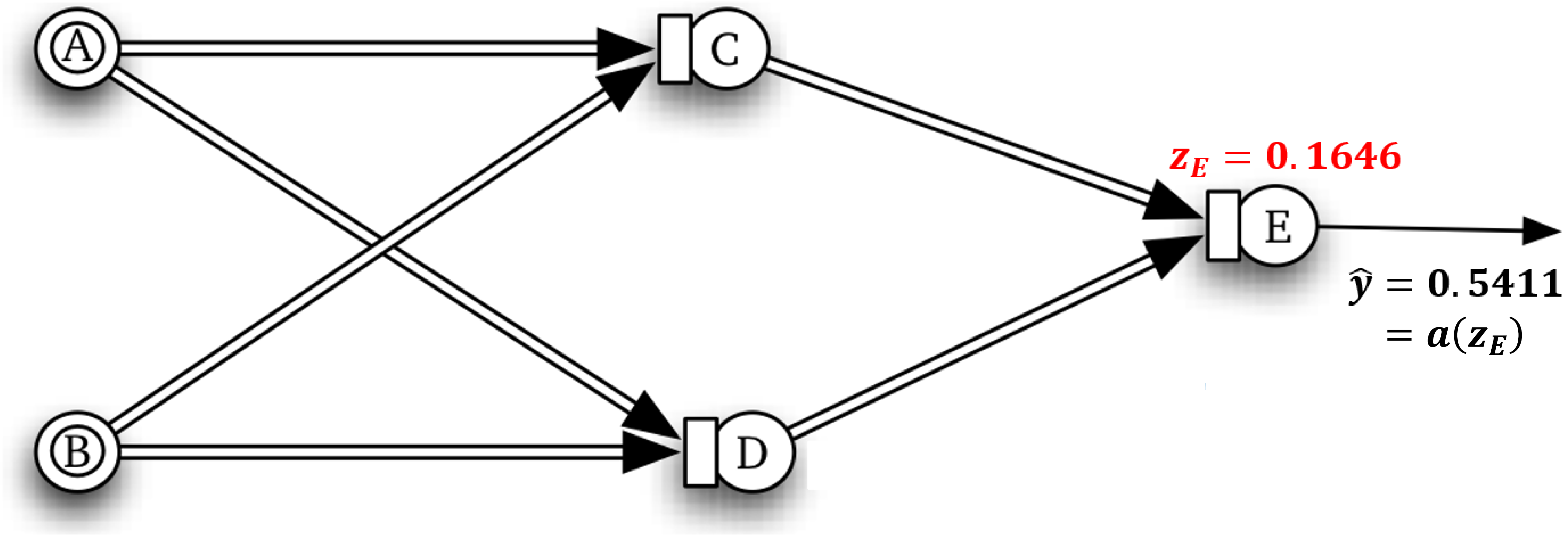
FEED FORWARD NETWORK



FEED FORWARD NETWORK



FEED FORWARD NETWORK



ANN IN MATRIX NOTATION – INPUT TO HIDDEN LAYER

Input: $\boldsymbol{X}_{(n\times p)} = \boldsymbol{X}_{(n\times 2)}$

$$\boldsymbol{X} = \begin{bmatrix} x_{A,1} & x_{B,1} \\ \vdots & \vdots \\ x_{A,n} & x_{B,n} \end{bmatrix}$$

Weights: $\boldsymbol{W}^{(1)}_{(p\times M)} = \boldsymbol{W}^{(1)}_{(2\times 2)}$

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} w_{AC} & w_{AD} \\ w_{BC} & w_{BD} \end{bmatrix}$$

Hidden units: $\boldsymbol{Z}^{(2)}_{(n\times M)} = \boldsymbol{Z}^{(2)}_{(n\times 2)}$

$$\boldsymbol{Z}^{(2)} = \begin{bmatrix} z_{C,1} & z_{D,1} \\ \vdots & \vdots \\ z_{C,n} & z_{D,n} \end{bmatrix} = \boldsymbol{X}\boldsymbol{W}^{(1)}$$

Activation function: $\boldsymbol{a}^{(2)}_{(n\times M)} = \boldsymbol{a}^{(2)}_{(n\times 2)}$

$$\boldsymbol{a}^{(2)} = \begin{bmatrix} 1/\left(1+e^{-z_{C,1}}\right) & 1/\left(1+e^{-z_{D,1}}\right) \\ \vdots & \vdots \\ 1/\left(1+e^{-z_{C,n}}\right) & 1/\left(1+e^{-z_{D,n}}\right) \end{bmatrix} = g\left(\boldsymbol{Z}^{(2)}\right)$$

ANN IN MATRIX NOTATION – HIDDEN TO OUTPUT LAYER

Activation function: $\boldsymbol{a}^{(2)}_{(n\times M)} = \boldsymbol{a}^{(2)}_{(n\times 2)}$

$$\boldsymbol{a}^{(2)} = g\left(\boldsymbol{Z}^{(2)}\right)$$

Weights: $\boldsymbol{W}^{(2)}_{(M\times K)} = \boldsymbol{W}^{(2)}_{(2\times 1)}$

$$\boldsymbol{W}^{(2)} = [\boldsymbol{w}_{CE} \quad \boldsymbol{w}_{DE}]$$

Output units: $\boldsymbol{Z}^{(3)}_{(n\times K)} = \boldsymbol{Z}^{(3)}_{(n\times 1)}$

$$\boldsymbol{Z}^{(3)} = \begin{bmatrix} \boldsymbol{Z}_{E,1} \\ \vdots \\ \boldsymbol{Z}_{E,n} \end{bmatrix} = \boldsymbol{a}^{(2)}\boldsymbol{W}^{(2)}$$

Activation function: $\boldsymbol{a}^{(2)}_{(n\times K)} = \boldsymbol{a}^{(2)}_{(n\times 1)}$

$$\hat{\boldsymbol{y}} = \boldsymbol{a}^{(3)} = \begin{bmatrix} 1/\left(1+e^{-\boldsymbol{Z}_{E,1}}\right) \\ \vdots \\ 1/\left(1+e^{-\boldsymbol{Z}_{E,n}}\right) \end{bmatrix} = g\left(\boldsymbol{Z}^{(3)}\right)$$

ANN IN MATRIX NOTATION

This *vanilla* neural net example can be expressed as:

$$\hat{y} = \boldsymbol{a}^{(3)} = g\left(\boldsymbol{Z}^{(3)}\right) = g\left[\boldsymbol{a}^{(2)}\boldsymbol{W}^{(2)}\right] = g\left[g\left(\boldsymbol{X}\boldsymbol{W}^{(1)}\right)\boldsymbol{W}^{(2)}\right]$$

In a nutshell, at each node, the neural net

1. computes a **weighted** **sum** of **inputs**
2. applies **activation** functions, and
3. sends a **signal**,

until the signal reaches the final **output** node.

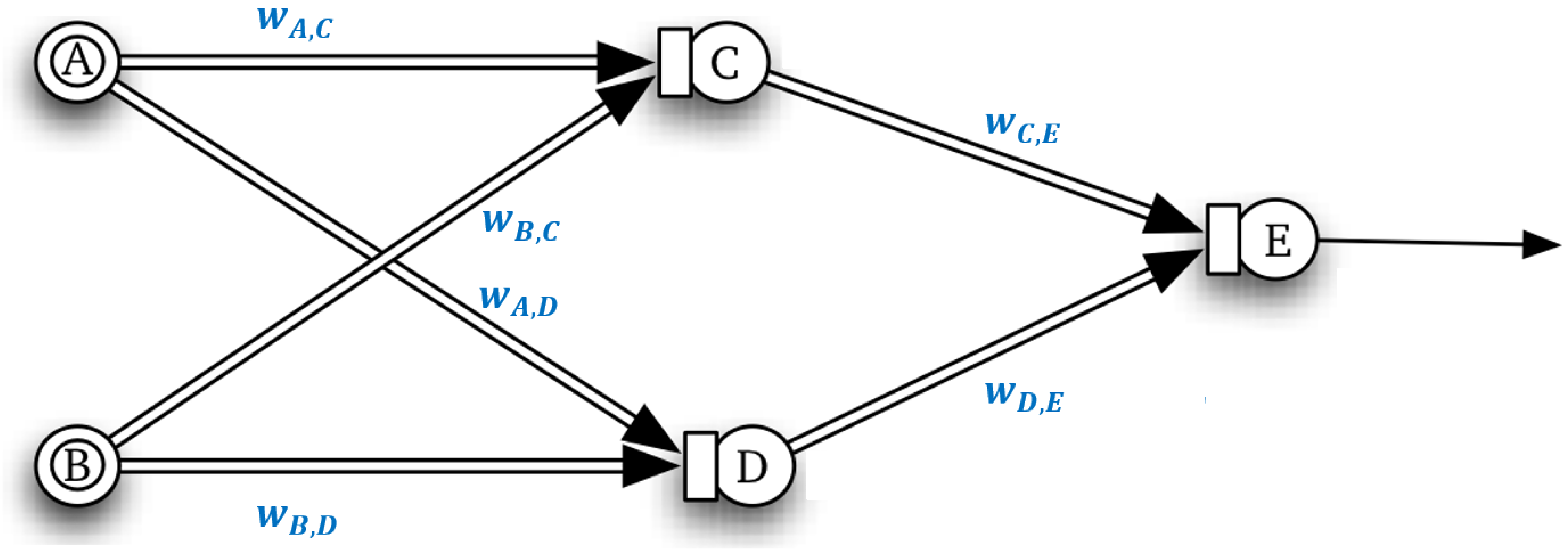
BACKPROPAGATION – ANN TRAINING

Given a signal, an ANN can produce an output, as long as the weights are specified.

For **supervised** learning tasks (i.e. when an ANN attempts to emulate the results of training examples), simply picking weights at random is a failing proposition.

Backpropagation is a method to optimize the choice of the weights against an error function $R(\mathbf{W})$.

BACKPROPAGATION OBJECTIVE



BACKPROPAGATION – ERROR FUNCTION

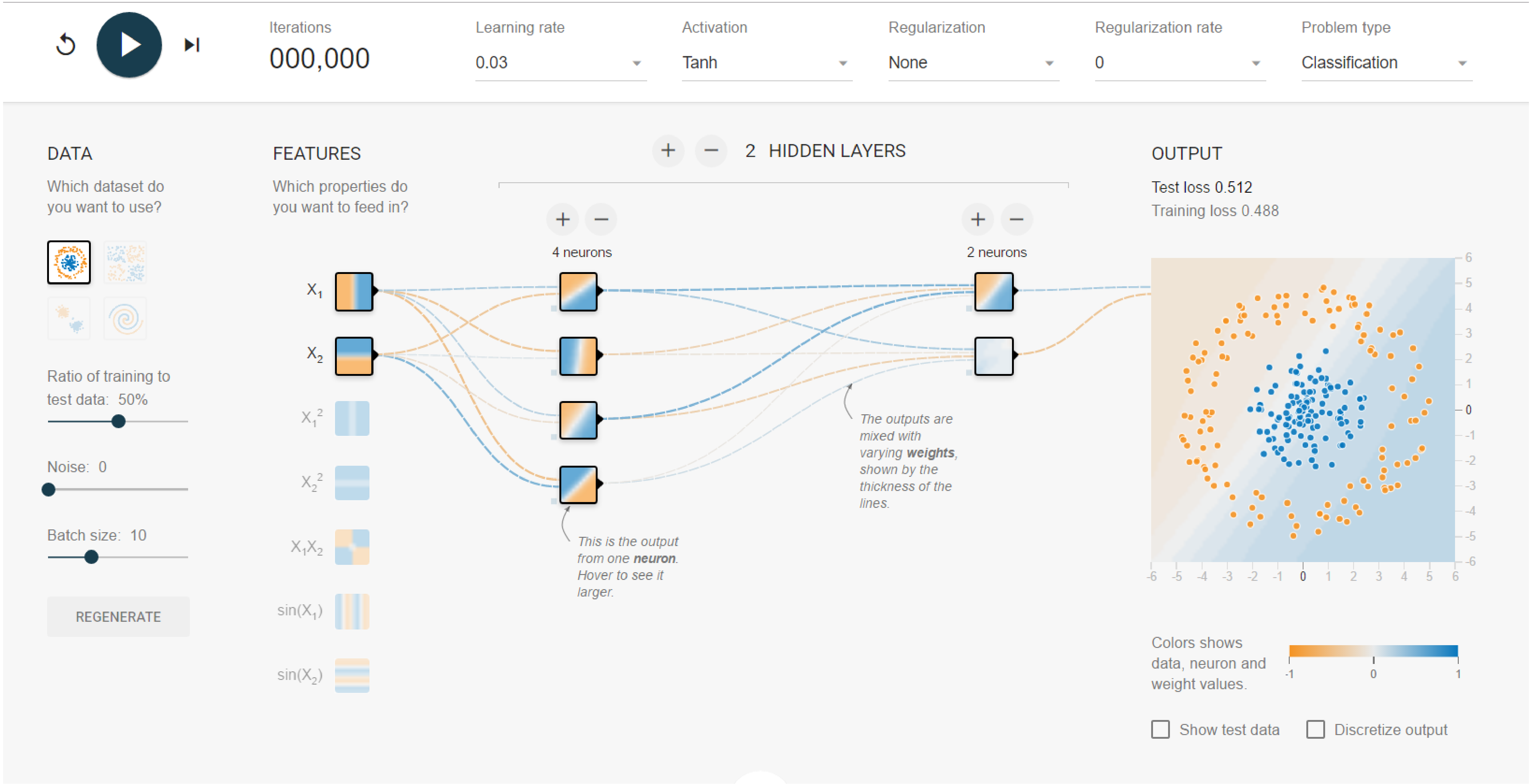
In regression problems, the **sum of squared errors** (SSE) is often used as the error function:

$$R(\boldsymbol{W}) = \sum_{i=1}^{n} \sum_{k=1}^{K} (\hat{y}_{ik}(\boldsymbol{W}) - y_{ik})^2$$

In classification problems, **cross-entropy** can be used:

$$R(\boldsymbol{W}) = -\sum_{i=1}^{n} \sum_{k=1}^{K} y_{ik} \ln[\hat{y}_{ik}(\boldsymbol{W})]$$

Either way, we want to minimize $R(\boldsymbol{W})$ with respect to $\boldsymbol{W}$ (**gradient descent**).



A NEURAL NETWORK PLAYGROUND

[D. Smilkov, S. Carter, TensorFlow.org]

STRENGTHS

ANNs can be quite **accurate** when making predictions – better than other algorithms with a proper set up.

ANNs often work when other things fail:

- when the relationship between attributes is **complex**
- when there are a lot of dependencies/**nonlinear relationships**
- **messy**, highly connected inputs (images, text and speech)
- non-linear classification

ANNs are relatively easy to set up (with available packages).

ANNs degrade gracefully (important in robotics).

LIMITATIONS

ANNs are relatively slow (creating and using) and prone to overfitting (may require **large/diverse** training set).

ANNs usually do not provide good interpretation (unlike decision trees or logistic regression, say). Can you live with that?

No algorithms for selecting the optimal network topology.

Even when they do perform better than other options, ANNs may not perform that much better due to **No Free-Lunch Theorems**; and they're susceptible to various forms of **adversarial attacks**.

DISCUSSION

The biggest challenge (to our minds) is to overcome the black box nature of ANNs. How important is it to you and your organization to be able to explain data-driven decisions?

NEURAL NETWORKS VIDEOS (BRILLIANT!)

1. Neural Networks Demystified, Welch Labs
   https://www.youtube.com/watch?v=bxe2T-V8XRs (first in the series)
2. Learning to See, Welch Labs
   https://www.youtube.com/watch?v=i8D90DkCLhI (first in the series)
3. Neural Networks, 3 Blue 1 Brown
   https://www.3blue1brown.com/videos/2017/10/9/neural-network

CASE STUDY: JAPANESE PHARMACY CHAIN

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

A Neural Network Application to Identify High-Value Customers for a Large Retail Store in Japan

(Ip, E., Johnson, J., Yada, K., Hamuro, Y., Katoh, N., Cheung, S. [2002], *Neural Networks in Business: Techniques and Applications*, Smith, K., Gupta, J. (eds), IRM Press)

CONTEXT

6 × as costly to sell to a new customer than to an existing one
(Kalakota, Robinson, Tapscott, 1999)

Annual customer retention up by 5% can lead to an 85% increase in profits
(Kalakota, Robinson, Tapscott, 1999)

Retaining the “right” customers plays a role in long-term profitability (Reicheld, 1993)

How can **loyal** and **profitable** customers be identified **early on**?

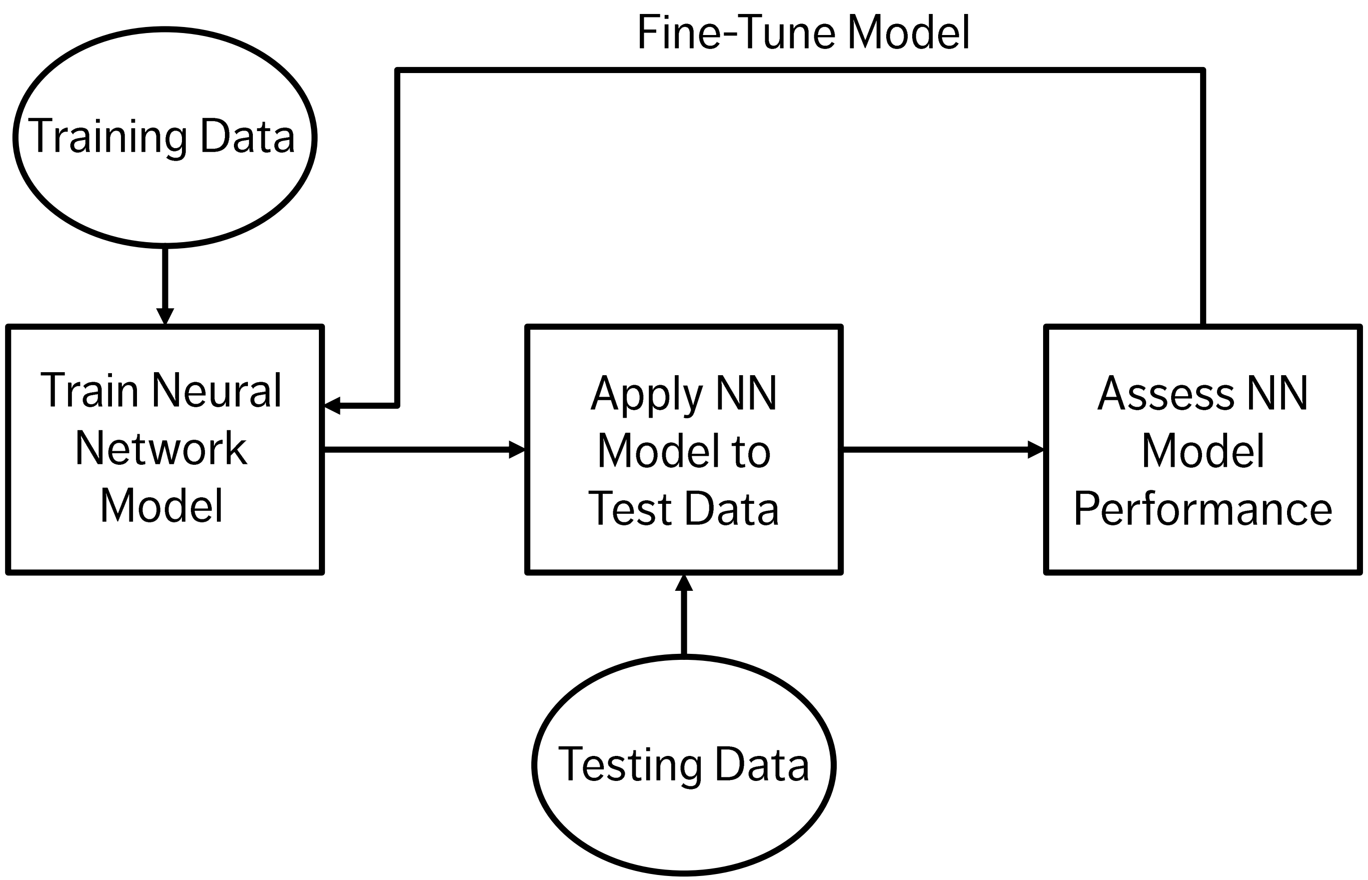
CONTEXT

Japanese drugstore chain *Pharma* (annual revenues ¥70B), 1300 retail stores, 2.3 million customers

Stores may use their own names, but operate under a central information system, which collects transaction and customer information since the early 1990s

Sales campaign information, customized coupons, free samples mailed to target group (i.e. loyal, high-value customers)

NEURAL NETWORK LEARNING FLOW



DATA

114,069 customers who made purchases during a 1-yr period

Customer value measured using

- frequency of visit (1 – 5 scale)
- profitability per visit (1 – 5 scale)

High-value customers (HVC): (4,5), (5,4), (5,5), making up 10.6% of observations

HVCs generate 52.5% of profits, 38.4% of revenues

DATA

Target variable: customer value

Input variables: total number of categories purchased, profit per visit, # of units purchased per visit, # visits, purchase of:

All variables were scaled from 0 – 1:

- paper product
- detergent
- eye drops
- kitchen cleaner
- bottled supplement
- hair care products
- fabric softener
- household cleaner
- toothpaste
- cold medicine

DATA

Learning set: 104,069 observations (selected randomly)

Training/Testing set ratio: 70-to-30

Validation set: remaining 10,000 observations

Calibration metrics: predictive accuracy, overall accuracy

predictive accuracy = # correctly predicted HVCs / # HVCs

overall accuracy = # correctly predicted class / # customers

RESULTS

Using a **Multilayered Feed Forward Neural Network** (MFFN), researchers were able to capture 80% of the HVCs by targeting (a model-specified) 25% of new customers.

At a threshold parameter value of 30%, the model performs 5 × better than randomly classifying customers.

Dataset	Training	Validation
Predictive Accuracy	55.6%	57.4%
Overall Accuracy	90.6%	91.2%
% Classified as HVC	10.6%	10.3%

DISCUSSION

Do you think the results of this case study are transferrable to other domains?

Would a similar MFFN provide the same level of predictive power in Canada?

DEEP LEARNING

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

"Humanity is on the verge of digital slavery at the hands of AI and biometric technologies. One way to prevent that is to develop inbuilt modules of deep feelings of love and compassion in the learning algorithms."

(A. Ray)

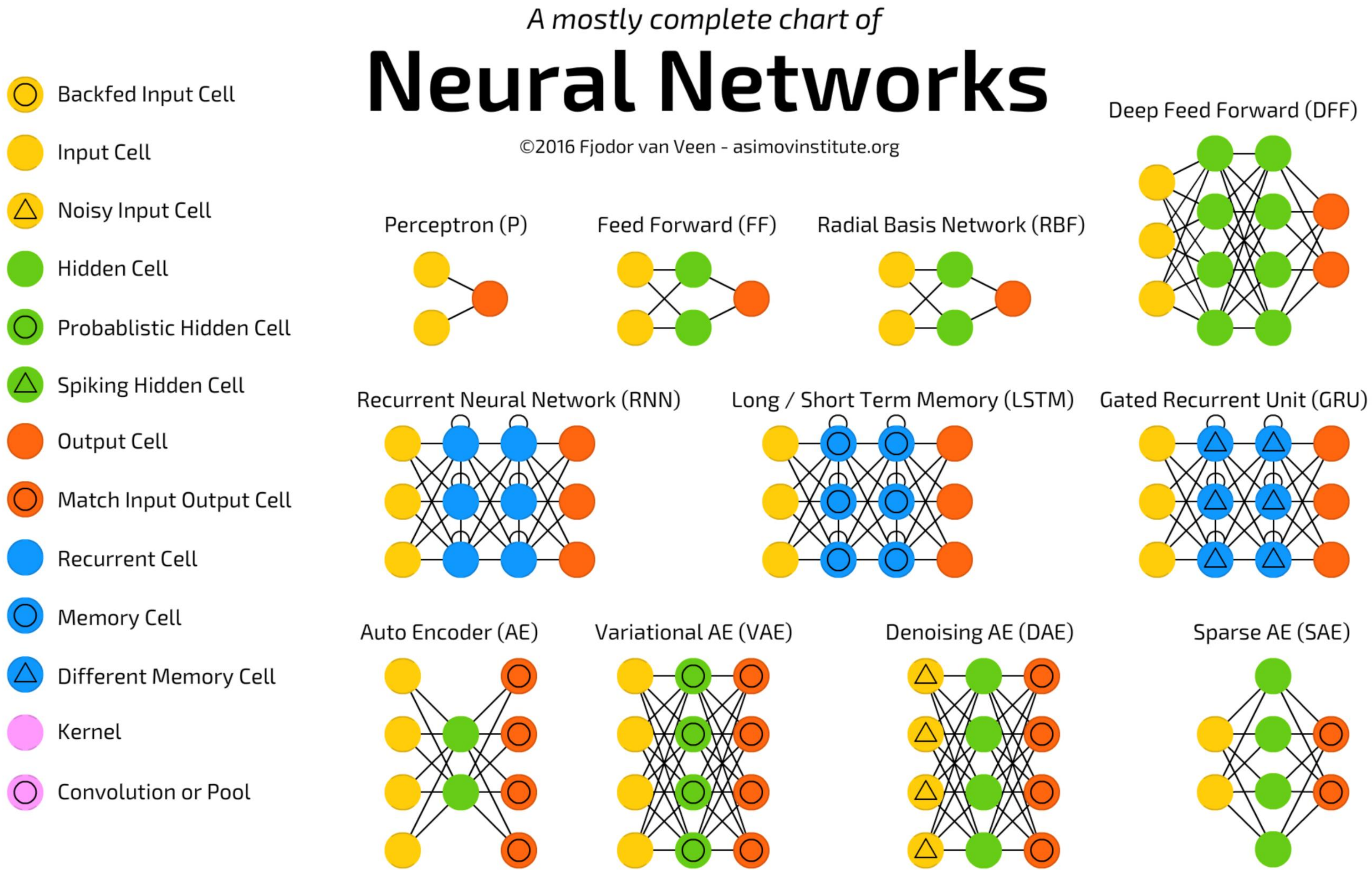
DEEP LEARNING NETWORKS

Deep Learning networks are simply ANNs with **a large number of hidden layers** (and various types of nodes)

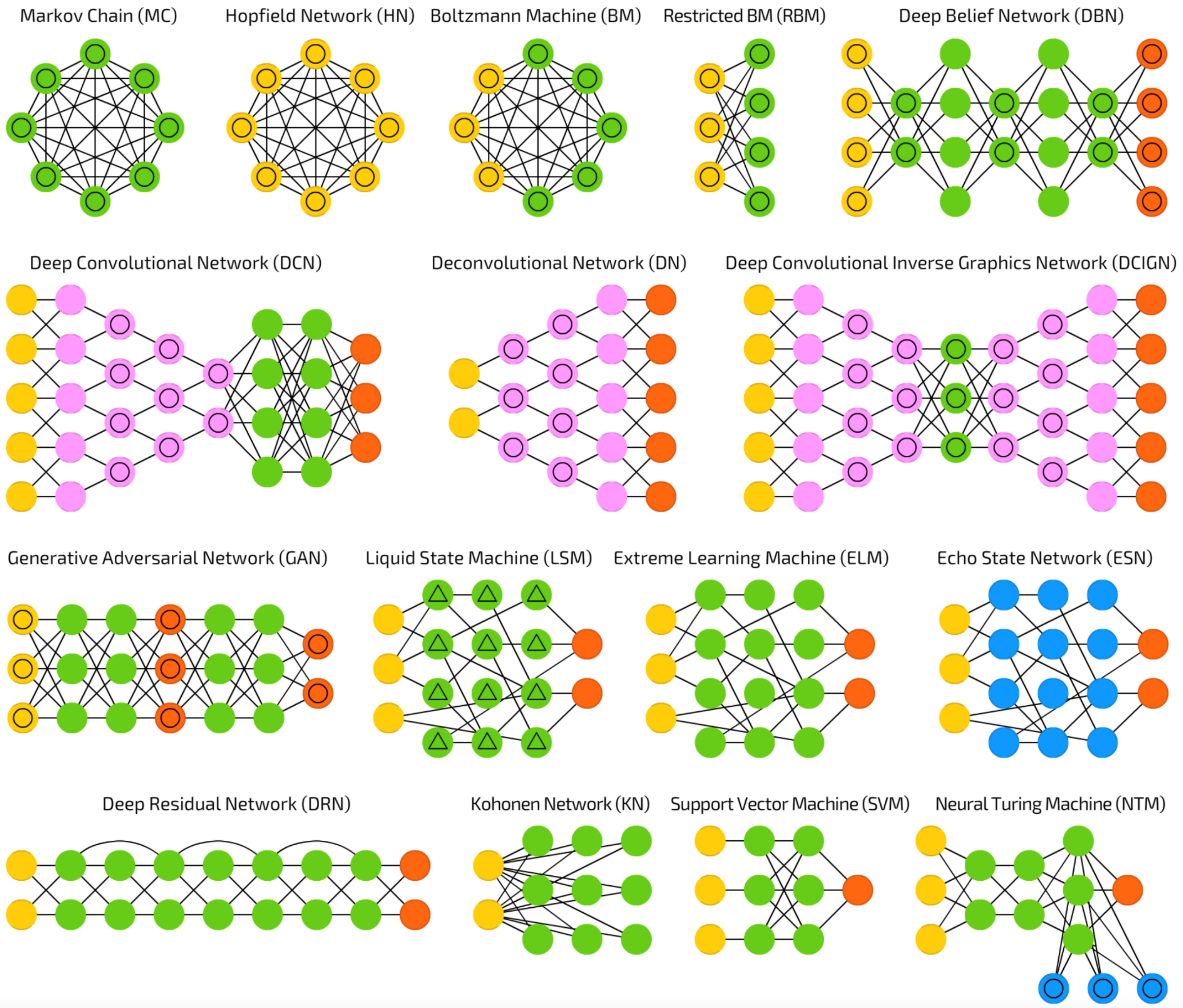
Types:

- *Convolution Neural Networks*
  Handwritten digit recognition, 99.7% accuracy in 2013, Self-driving cars
- *Recurrent Neural Networks*
  Natural language processing (speech recognition, machine translation, etc.)
- *Autoencoders*
- *Restricted Boltzmann Machines*
  BellKor's Pragmatic Chaos, Netflix Prize, 2009

[Fjodor van Veen, Asimov Institute, 2016]



[Fjodor van Veen, Asimov Institute, 2016]



LIMITATIONS

Require **large**, **diverse**, and **correctly labeled** training sets.

Accurate on average, but they can still be **spectacularly** wrong.

They can be "hacked" (NFL).

Humans don't need that much labeled data to make decisions: so **what's really going on under the hood**? (3rd AI Winter?)

LIMITATIONS

Require **large**, **diverse**, and **correctly labeled** training sets.

Accurate on average, but they can still be **spectacularly** wrong.

They can be "hacked" (NFL).

Humans don't need that much labeled data to make decisions: so **what's really going on under the hood**? (3rd AI Winter?)

DISCUSSION

How do we align autonomous A.I. goals with human/cultural values?

SUPPLEMENTAL MATERIAL

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

HISTORICAL TIMELINE

First AI Winter (1974 – 1980)

- High expectations but next to nothing to show for them
- Lack of computing power, funding withdrawn

Revival

- Japan's Fifth Generation Project (1982 – 1992)
- Defense Advanced Research Projects Agency funding (1983)
- ***Backpropagation*** (1985)
- 1st IEEE Conference (1987)

HISTORICAL TIMELINE

Second AI Winter (1987 – 1993)

- Expensive to maintain; massive funding cuts
- Bad learners, not great with uncertainty
- Japan’s FGP goals not met

Modern AI (1993 – 2001)

- *Huge* leap in computing power
- Deep Blue and “intelligent agents” paradigm

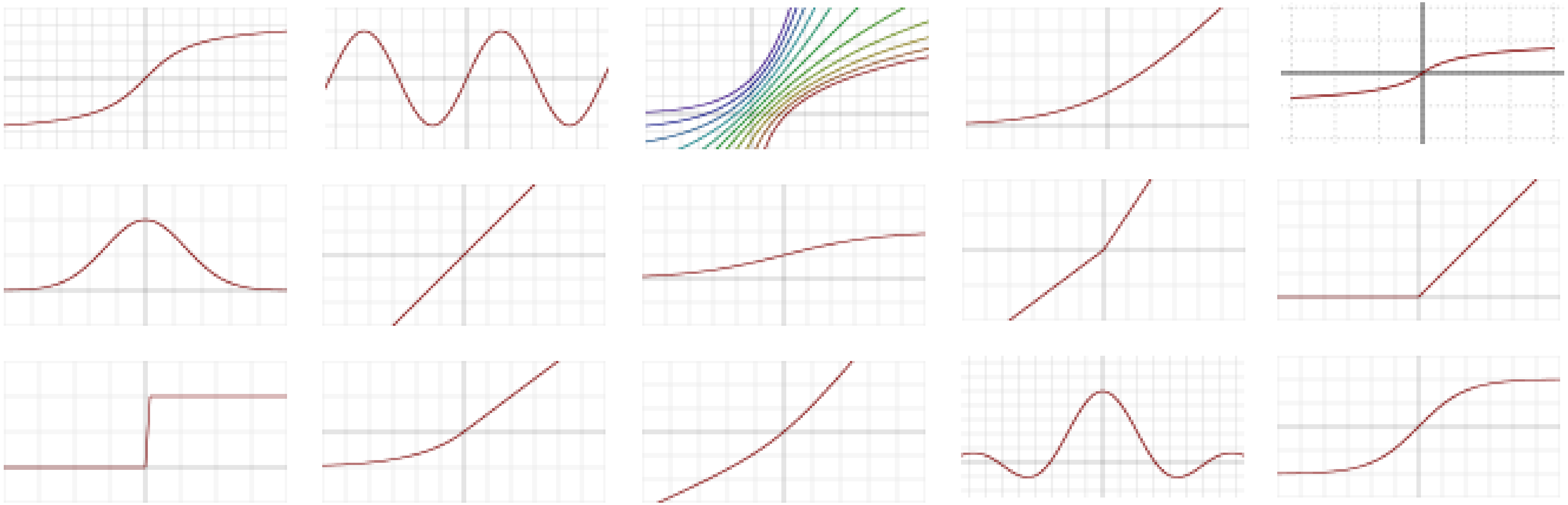
HISTORICAL TIMELINE

Deep Learning and Big Data (2001 – present)

- Parallel computing power
- Terabytes of data
- Deep neural networks
- Human Brain Project (10yr project, 1.3B USD, 2013 – 2023)
- Self-driving vehicles?

[https://en.wikipedia.org/wiki/Activation_function]

ACTIVATION FUNCTIONS



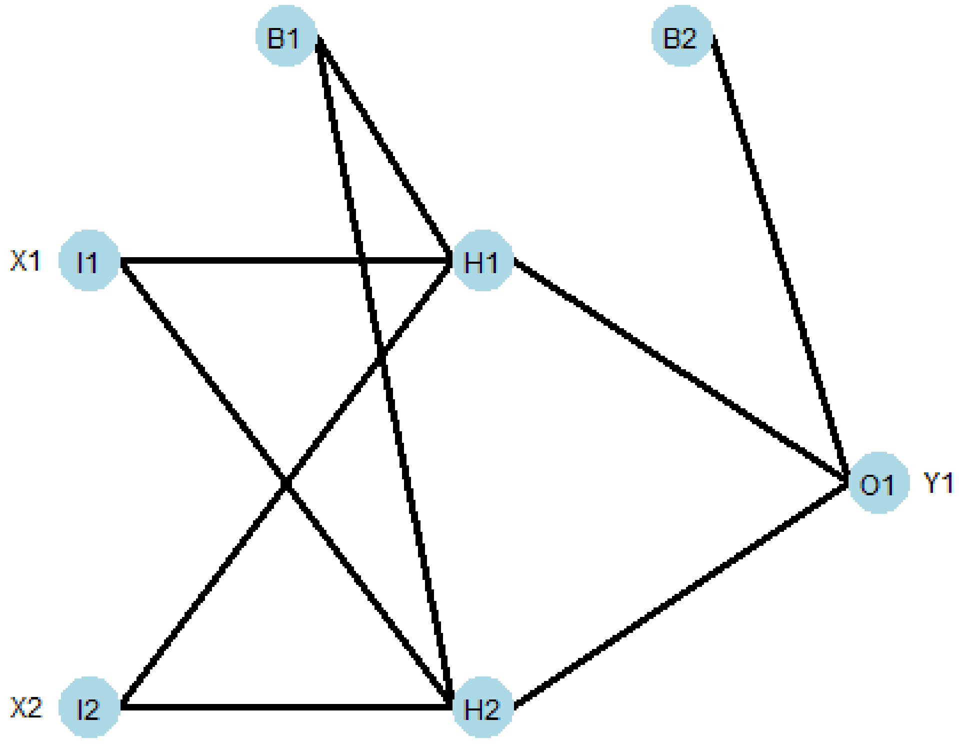
There are **many** options: what effect(s) does the choice have, if any?

BIAS

A **bias** term $\boldsymbol{b}$ can be introduced to all units in hidden and output layers, which act as intercepts for $\boldsymbol{Z}$.

$$\boldsymbol{Z}^{(2)} = \begin{bmatrix} \boldsymbol{b}_C + \boldsymbol{Z}_{C,1} & \boldsymbol{b}_D + \boldsymbol{Z}_{D,1} \\ \vdots & \vdots \\ \boldsymbol{b}_C + \boldsymbol{Z}_{C,n} & \boldsymbol{b}_D + \boldsymbol{Z}_{D,n} \end{bmatrix} \text{ and}$$

$$\boldsymbol{Z}^{(3)} = \begin{bmatrix} \boldsymbol{b}_E + \boldsymbol{Z}_{E,1} \\ \vdots \\ \boldsymbol{b}_E + \boldsymbol{Z}_{E,n} \end{bmatrix}$$



[https://archive.ics.uci.edu/ml/datasets/wine]

EXAMPLE – WINES DATA

Sample ($N = 178$) of wines grown in the same region of Italy, which all come from three different cultivars (**target**).

$p = 13$ chemical and non-chemical properties (**features**) are recorded.

Variables:

- alcohol
- malic acid
- ash
- alkalinity of ash
- magnesium
- total phenols
- flavonoids
- non-flavonoid phenols
- colour intensity
- hue
- OD280/OD315
- proline

https://www.data-action-lab.com/wp-content/uploads/2019/03/Advanced.zip

REFERENCES

ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

REFERENCES

Do Androids Dream of Electric Sheep?, the Guardian.

Neural Networks Demystified, Welch Labs.

Lawrence, S., Giles, C.L., Tsoi, A.C. (1996), "What size neural network gives optimal generalization? Convergence properties of backpropagation". Technical Report UMIACS-TR-96-22 and CS-TR-3617, Institute for Advanced Computer Studies, University of Maryland, College Park.

Elissee, A., Paugam-Moisy, H. (1997), "Size of multilayer networks for exact learning: analytic approach". Advances in Neural Information Processing Systems 9, Cambridge, MA: The MIT Press, pp.162-168.

Y. Bengio, Y. LeCun (2007), Scaling learning algorithms towards AI, in L. Bottou, O. Chapelle, D. DeCoste, J. Weston (eds), Large-Scale Kernel Machines, MIT Press

A Beginner's Guide to Understanding Convolutional Neural Networks (part 1, part 2)

Fitting a neural network in R

Training and Visualizing a Neural Network With R

REFERENCES

A Neural Network Playground

Neural Networks and Deep Learning

A Beginner's Guide to Neural Networks With R!

Neural Networks, Manifolds, and Topology

Wikipedia entry for ANNs: https://en.wikipedia.org/wiki/Artificial_neural_network

C. Stergiou, D. Siganos, Neural Networks

Artificial Intelligence – Neural Networks

UFLDL Tutorial, Stanford University

The Neural Network Zoo, Asimov Institute

https://en.wikipedia.org/wiki/Reinforcement_learning

REFERENCES

http://www.scmp.com/tech/enterprises/article/2095929/alphago-vanquishes-worlds-top-go-player-marking-ais-superiority

https://www.theguardian.com/technology/2017/jan/30/libratus-poker-artificial-intelligence-professional-human-players-competition

http://www.cbc.ca/news/canada/newfoundland-labrador/artificial-intelligence-robots-battle-memorial-university-starcraft-1.4298844

https://www.digitaltrends.com/cool-tech/japanese-ai-writes-novel-passes-first-round-nationanl-literary-prize/

https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer/

https://techcrunch.com/2016/09/01/this-trailer-made-by-ibms-watson-proves-ai-knows-how-to-creep-you-out/

http://www.bbc.com/news/health-36482333

https://www.digitaltrends.com/computing/cdc-flu-forecasting/

REFERENCES

https://www.newscientist.com/article/2147472-ai-spots-alzheimers-brain-changes-years-before-symptoms-emerge/

Explaining and harnessing adversarial examples, I.J. Goodfellow, J. Shlens, C. Szegedy, ICLR 2015, https://arxiv.org/abs/1412.6572

https://arxiv.org/pdf/1710.08864.pdf

http://blog.ycombinator.com/how-adversarial-attacks-work/?imm_mid=0f81cc&cmp=em-data-na-na-newsltr_20171115

http://www.kurzweilai.net/

https://www.theguardian.com/technology/2014/jun/08/super-computer-simulates-13-year-old-boy-passes-turing-test

http://www.foxnews.com/tech/2016/03/25/microsoft-takes-tay-chatbot-offline-after-trolls-make-it-spew-offensive-comments.html

REFERENCES

http://www.dailymail.co.uk/sciencetech/article-4747914/Facebook-shuts-chatbots-make-language.html

http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

https://www.quora.com/What-is-the-Manifold-Hypothesis-in-Deep-Learning

No free lunch theorems for optimization, D.H. Wolpert, W.G. Macready, IEEE Transactions on Evolutionary Computation 1, 1997

https://www.inverse.com/article/29879-knock-knock-joke-humor-ai-neural-network

https://en.wikipedia.org/wiki/DARPA_Grand_Challenge (2005)

https://www.cnbc.com/2017/09/12/apple-unveils-a11-bionic-neural-engine-ai-chip-in-iphone-x.html

https://www.gizmodo.com.au/2015/09/this-neural-networks-hilariously-badimage-descriptions-are-still-advanced-ai/

REFERENCES

https://www.gizmodo.com.au/2015/07/googles-neural-network-chatbot-can-discuss-philosophy-and-it-troubles/

https://www.gizmodo.com.au/2015/08/this-neural-network-is-hilariously-bad-at-describing-outer-space/

https://splinternews.com/these-bots-wrote-and-delivered-their-own-ted-talks-sor-1793848398

Universal adversarial perturbations, Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, Pascal Frossard, Oct 26, 2016, https://arxiv.org/abs/1610.08401

https://www.cnet.com/news/elonmusk-artificial-intelligence-world-wariii-russia-china/

https://www.technologyreview.com/s/546301/will-machines-eliminate-us/

https://www.technologyreview.com/s/609070/inside-the-moonshot-effort-to-finally-figure-out-the-brain/

Is AI riding a one-trick pony? MIT Technology Review, Sep 29, 2017, https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/

REFERENCES

Axios, Sep 15, 2017, https://www.axios.com/ai-pioneer-advocates-starting-over-2485537027.html

TechCrunch, Sep 30, 2017, https://techcrunch.com/2017/09/30/ai-hype-has-peaked-so-whats-next

Fortune, Jun 28, 2017, http://fortune.com/2017/06/28/ibm-watson-ai-healthcare/

https://stuff.mit.edu/people/mitter/publications/121_Testing_Manifold.pdf

https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

http://robots.stanford.edu/papers/thrun.stanley05.pdf

https://medium.com/@samim/generating-captions-c31f00e8396e

R. Kolakota, M. Robinson, D. Tapscott [1999], *E-business 2.0: Roadmap for Success*, Addison-Wesley

F.F. Reichheld [1993], *Loyalty-based management*, Harvard Review, March/April, pp.64-73

https://distill.pub/2017/aia/

https://distill.pub/2018/building-blocks/