How Important Is Weight Symmetry in Backpropagation?

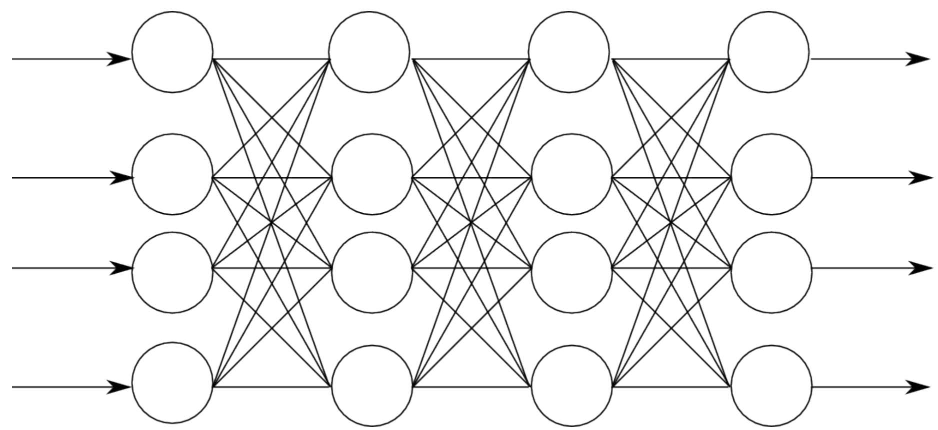
Qianli Liao, Joel Z. Leibo, Tomaso Poggio

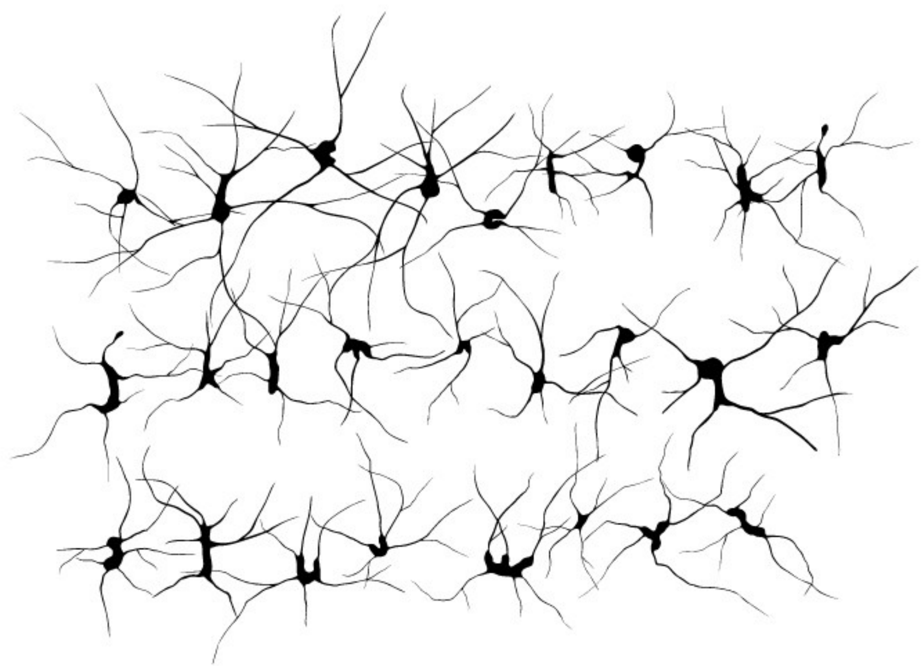
Center for Brains, Minds and Machines
McGovern Institute, MIT

Center for Brains, Minds & Machines

Massachusetts Institute of Technology

Biological plausibility of deep learning





Biological plausibility of deep learning

- Important question for both neuroscience and machine learning
- Has been heavily debated since 1980s

Biological plausibility of deep learning

Since most successful deep learning models are trained with Backpropagation,

It becomes a priority to answer:

Is backpropagation biologically plausible?

– Many past studies concluded NO

– We tend to say YES based on recent findings.

Why is backpropagation interesting?

- Deep Feedforward Neural Networks (DNNs):
  a key model of deep learning,
  successful in many domains
- DNNs: universal approximators
  --> good solutions exist!
- Surprise: good solutions are found
- How? By backpropagation (BP)

- Brain: arguably some form of DNN
- Could the brain implement backpropagation?

Could the brain implement BP?

- Many studies answered NO:
  Crick, F.1989;
  Mazzoni, Andersen, and Jordan 1991;
  O’Reilly 1996;
  Chinta and Tweed 2012;
  Bengio et al. 2015
  and many more …

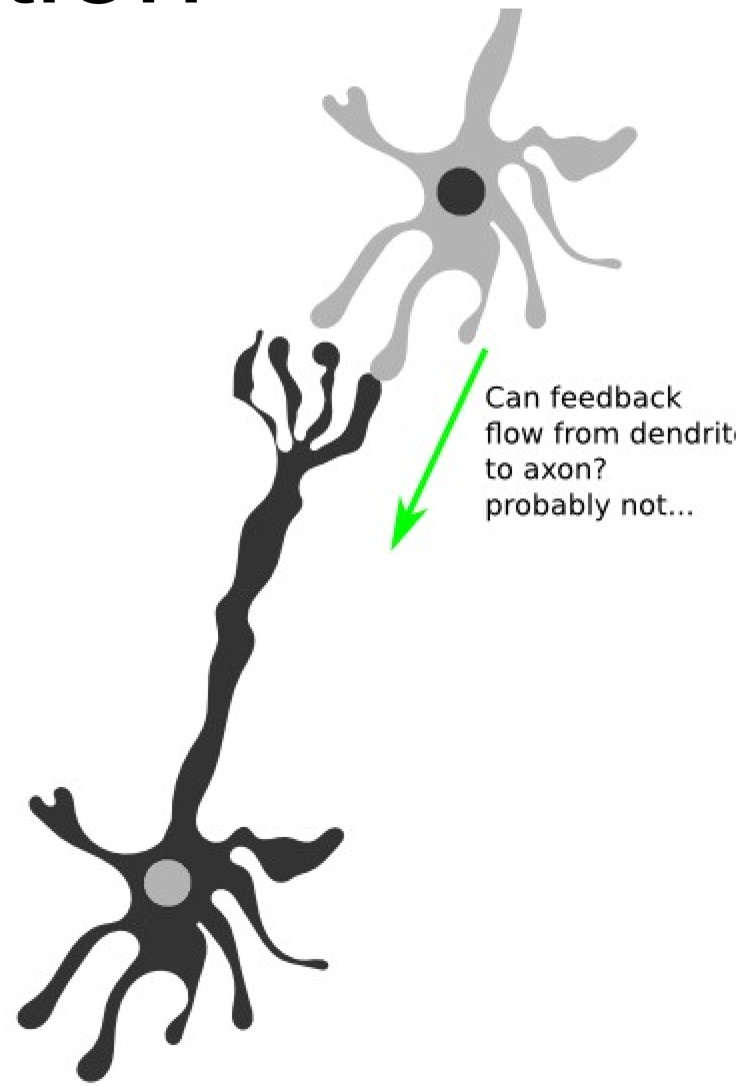
- Summary of the difficulties of implementing BP in the brain:
  (1) feedback weights = feedforward weights
  (2) forward and backward passes require different computations
  (3) error gradients must be stored separately from activations.

Biological difficulties of backpropagation

(1) feedback weights = feedforward weights (i.e., “weight transport problem”)

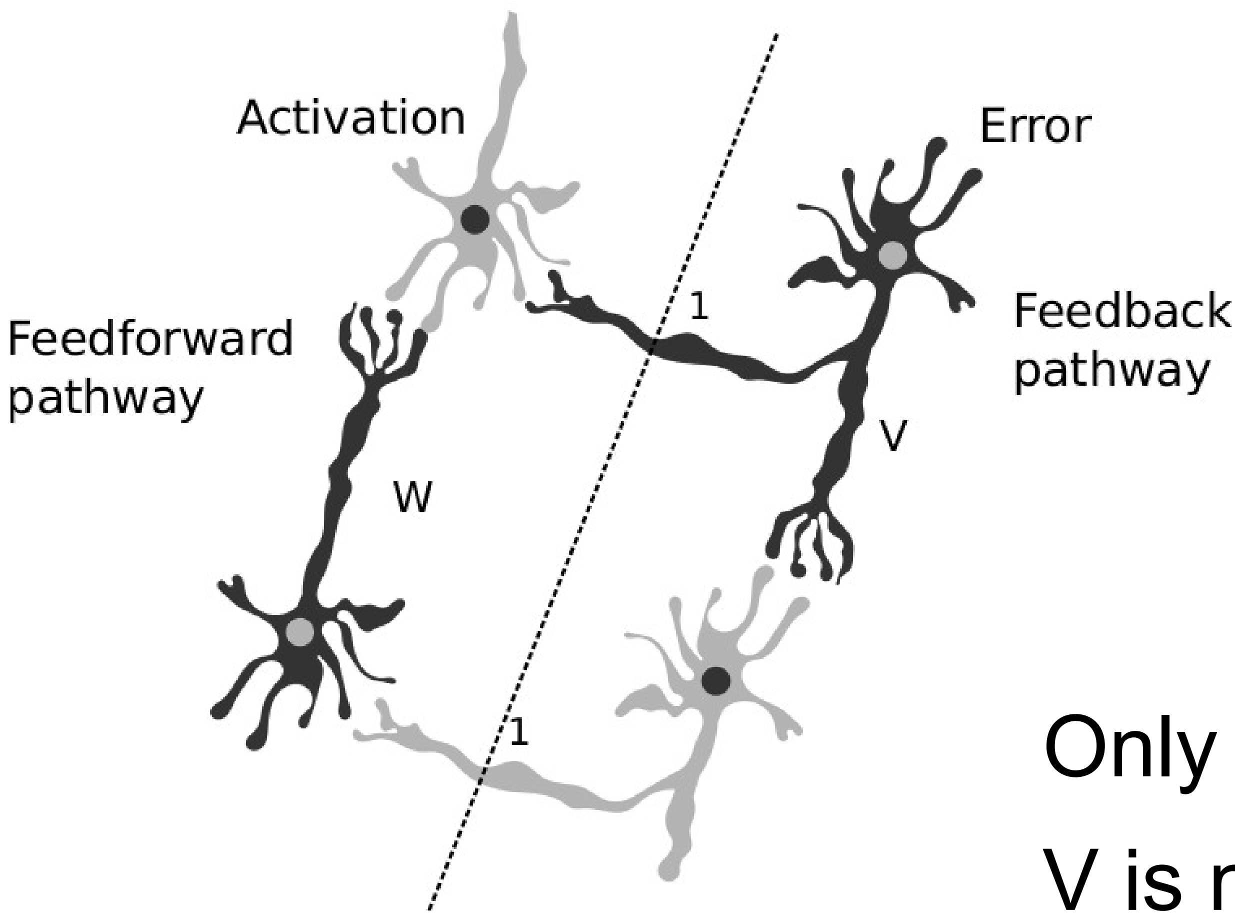
(2) forward and backward passes require different computations

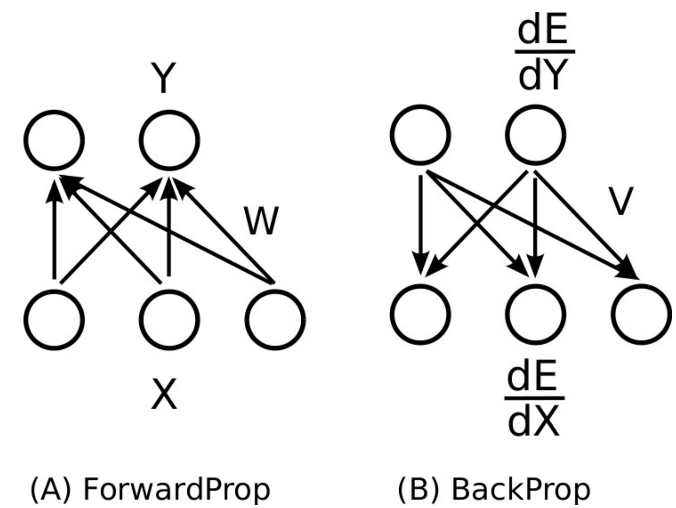
(3) error gradients must be stored separately from activations.



- Cannot be implemented by
  a pair of neurons (as show on the right).

Simple solution: error network



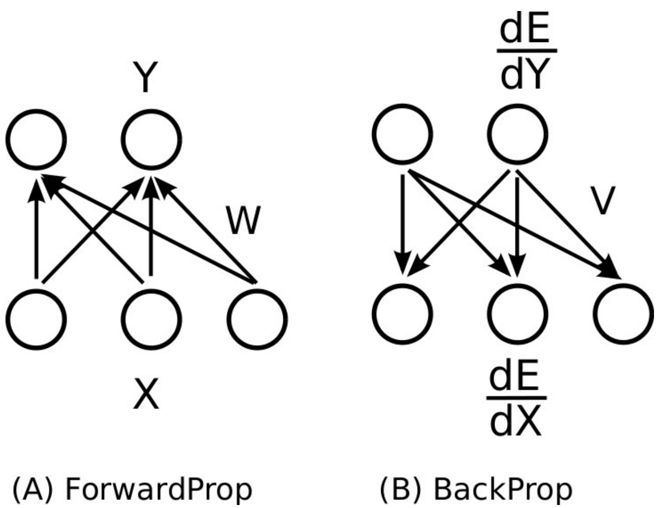


Only problem left:

V is not equal to W

We reduced 3 difficulties to 1!

Asymmetric backpropagation



Standard BP: V = W

Brain: V ≠ W

- **<u>Main finding!!:</u>**
  Even if V ≠ W,
  we can achieve
  strong performance
  (when combined
  with Batch Normalization
  and/or Batch Manhattan)
- **Good news for the brain!**

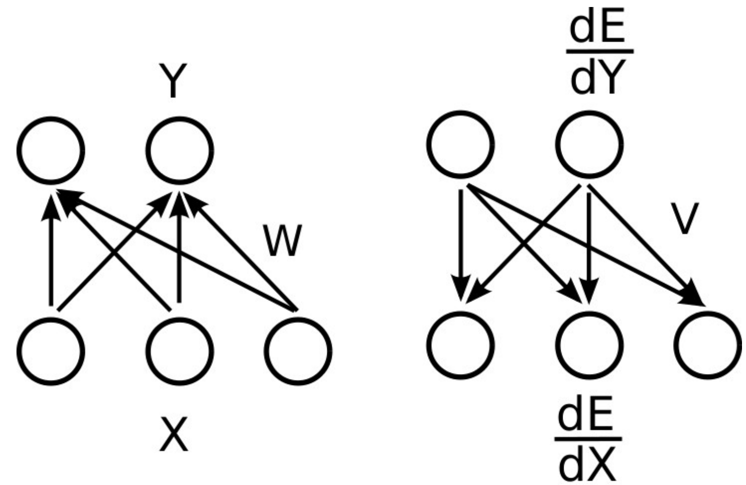
Main results

- Ideally V = W, but biologically implausible
- What if V = sign(W)?

  – vanishing or exploding gradients
- + Batch Normaliazation (Ioffe and Szegedy 2015)

  or Batch Manhattan

  – It works!
- Batch Manhattan = when updating the weights, use only the signs of the gradients (discard the magnitudes)



(A) ForwardProp (B) BackProp

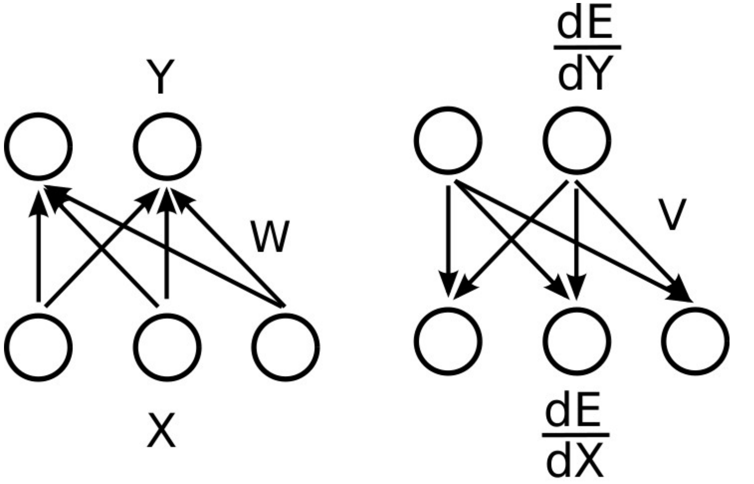
Main results

- What if

  V = M * sign(W)?

  where M is a random value from [0,1]

  – it works too!



Result:

The magnitudes seem to be NOT important

Main results

- What if

  V = M * sign(W) * Sp?

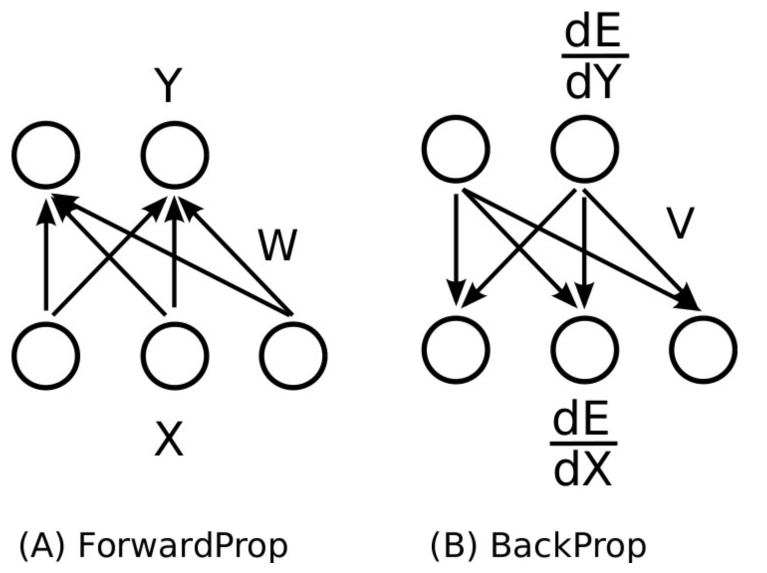
  where M and Sp are random.

  Sp = 1 or -1 with p and (1-p) probability

  Sp: sign concordance probability

  Result:

  The effectiveness of backpropagation is strongly dependent on high sign concordance probability.



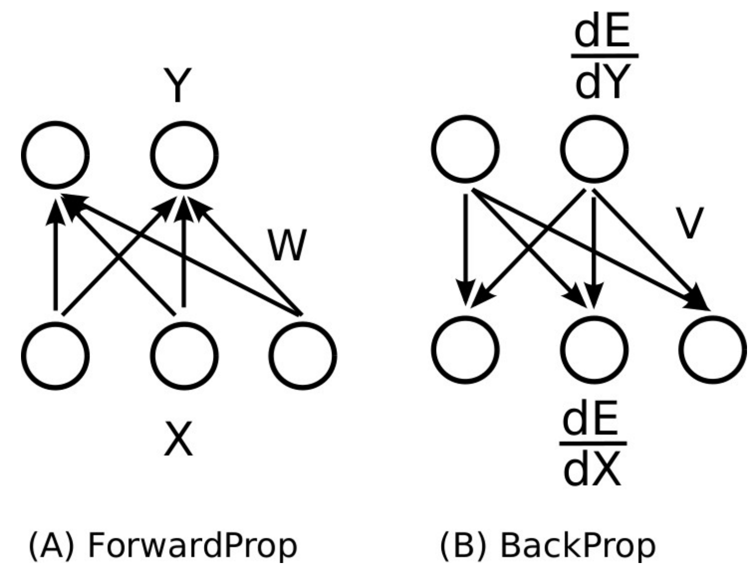
Main results

- What if

  V = fixed random

  – it outperforms chance, but not as good as standard SGD

  – batch normalization is still required



Related work: Other biologically plausible neural networks

Two camps:

- Bidirectionally connected networks:

  1. Recirculation Algorithms (Hinton and McClelland 1988)
  2. Generalized Recirculation Algorithms (O'Reilly 1996)

- BP Variants:

  1. Random feedback weights support learning in DNN. (Lillicrap et al. 2014)
  2. This work.

Related work:
Bidirectionally connected networks

- Pros

  Less assumptions.

  Weight symmetry is learned automatically

- Cons

  Stability

  Biological consistency (?)

  Performance

Compare to:
Random feedback weights support learning in DNN
(Lillicrap et al. 2014)

Our work:

- Found that **Batch Normalization** and **Batch Manhattan** are critical for the performance **(important)**
- Found that the **signs** of feedback are important
- Many more datasets
- Much deeper models

Contributions

- First **systematic** study of asymmetric forward and backward weights in BP
- We show that there exist variants of BP may be biologically implementable
- We identify the crucial role of **Batch Normalization** and **Batch Manhattan** in asymmetric BP.
- We demonstrate the importance of **signs**, instead of magnitudes of both feedback weights and gradient updates (i.e., Batch Manhattan update rule).

--- consistent with the notion that brain's algorithm is noise-tolerant.