Random feedback weights support learning in deep neural networks

Timothy P. Lillicrap[1*], Daniel Cownden[2], Douglas B. Tweed[3,4], Colin J. Akerman[1]

Gatsby Tea Talk

02/10/2015

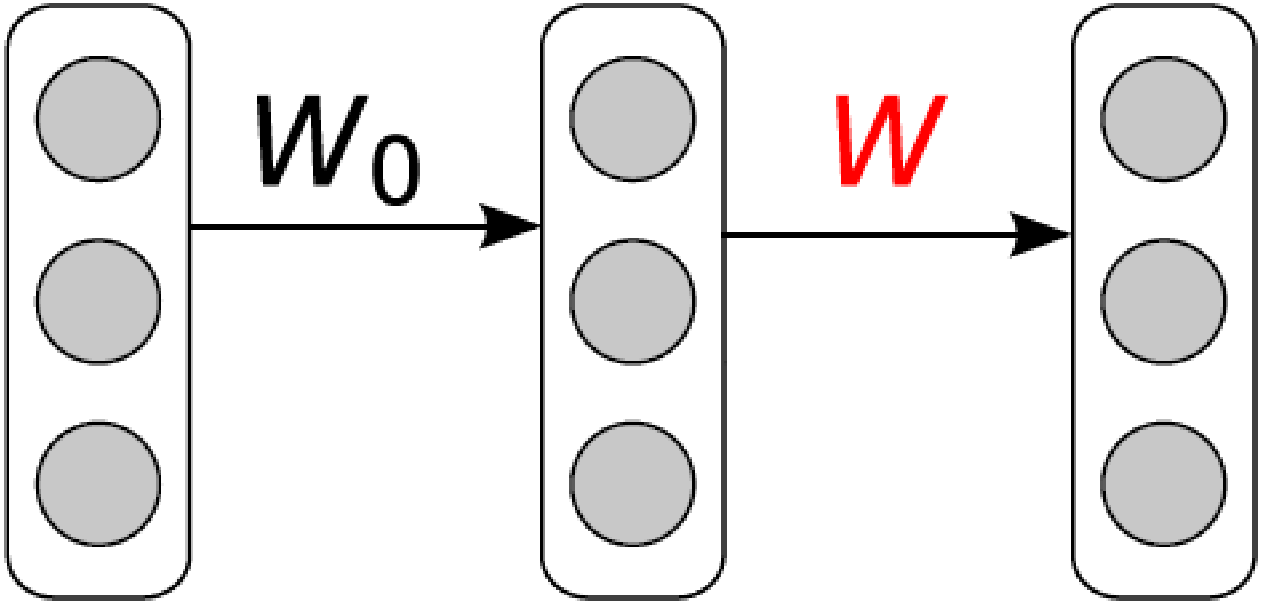
Presenter: Vincent Adam

What's in it

- A proposed 'biophysically realistic' alternative to back propagation

- Sketch of proofs of when it works

Feed Forward NN



x $\qquad h = W_0 x$ $\qquad y = Wh$

Supervised Learning

- data $\{x, \tilde{y}\}$
- error $\mathcal{L} = \frac{1}{2}||y - \tilde{y}||^2$

Back Prop

Output weights

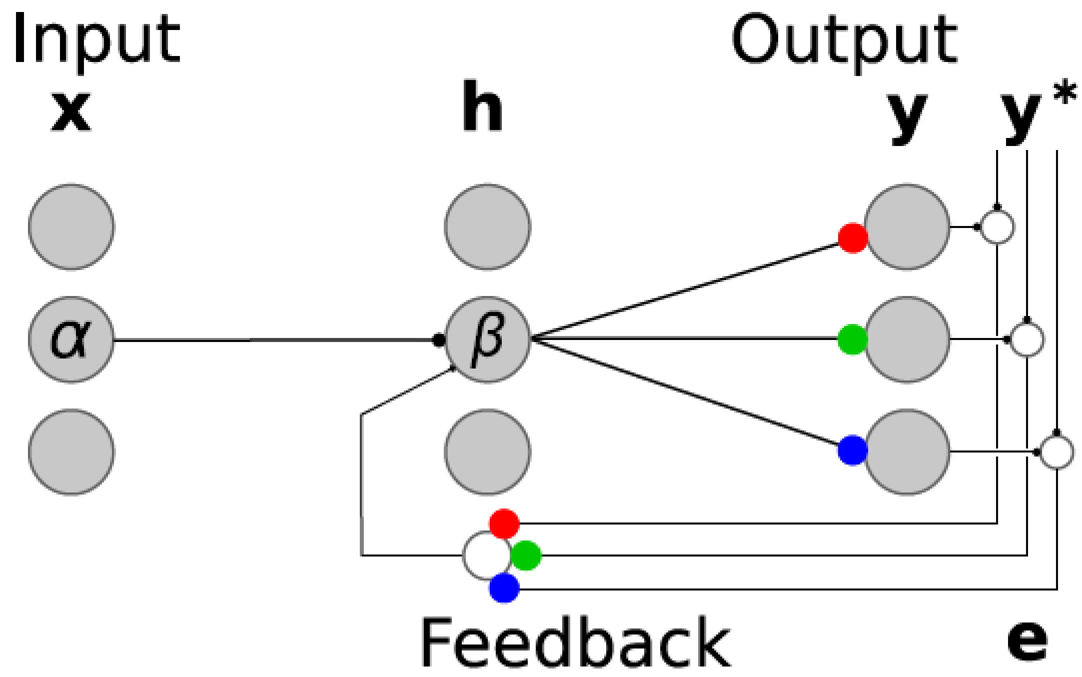
$$\Delta W \propto \frac{\partial \mathcal{L}}{\partial W} = -eh^T$$

Input weights

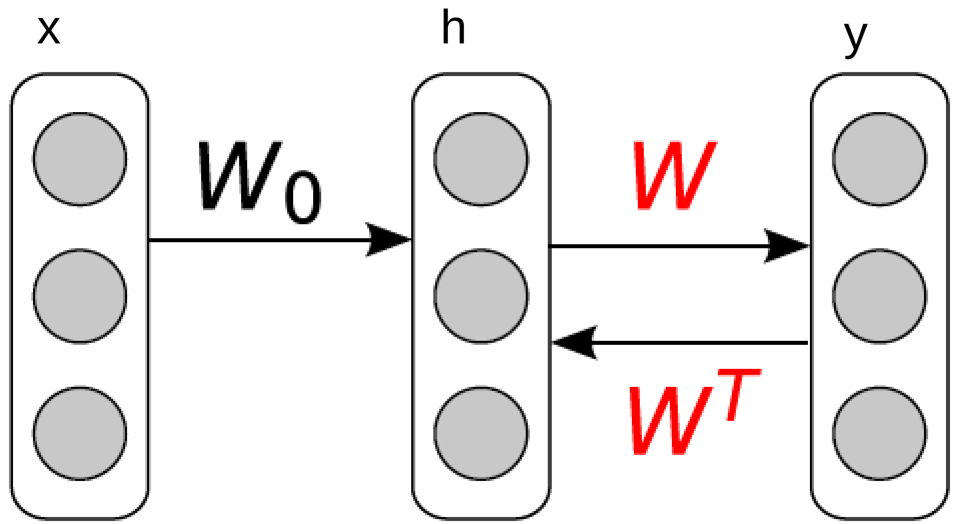
$$\Delta W_0 \propto \frac{\partial \mathcal{L}}{\partial W_0} = \frac{\partial \mathcal{L}}{\partial h}\frac{\partial h}{\partial W_0}$$

$$\propto -(W^T e)x^T$$

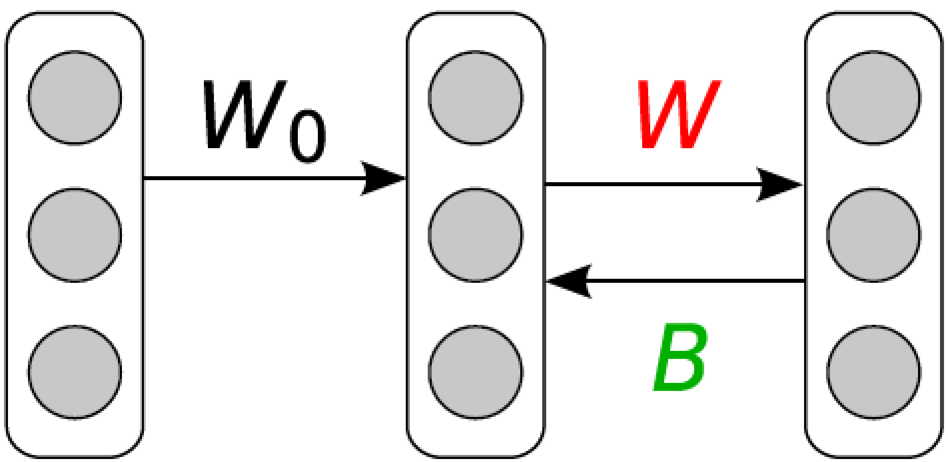
Transport of synaptic weight information



Feedback Alignment



$$\Delta W_0 \propto -(W^T e) x^T$$
$$\propto -\Delta h_{BP} x^T$$



$$\Delta W_0 \propto -(Be) x^T$$
$$\propto -\Delta h_{FA} x^T$$

Claim

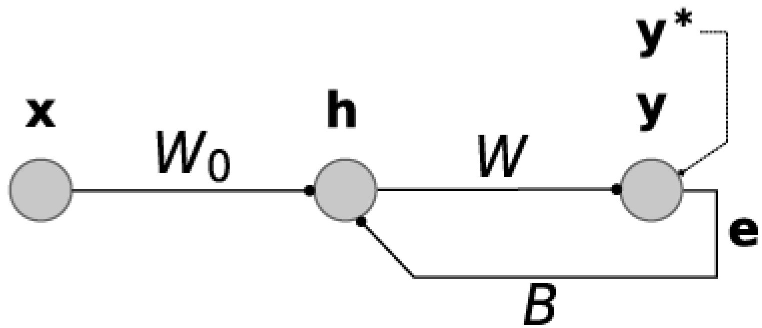
- B need not be exactly W^T

- It suffices that $e^T W B e > 0$

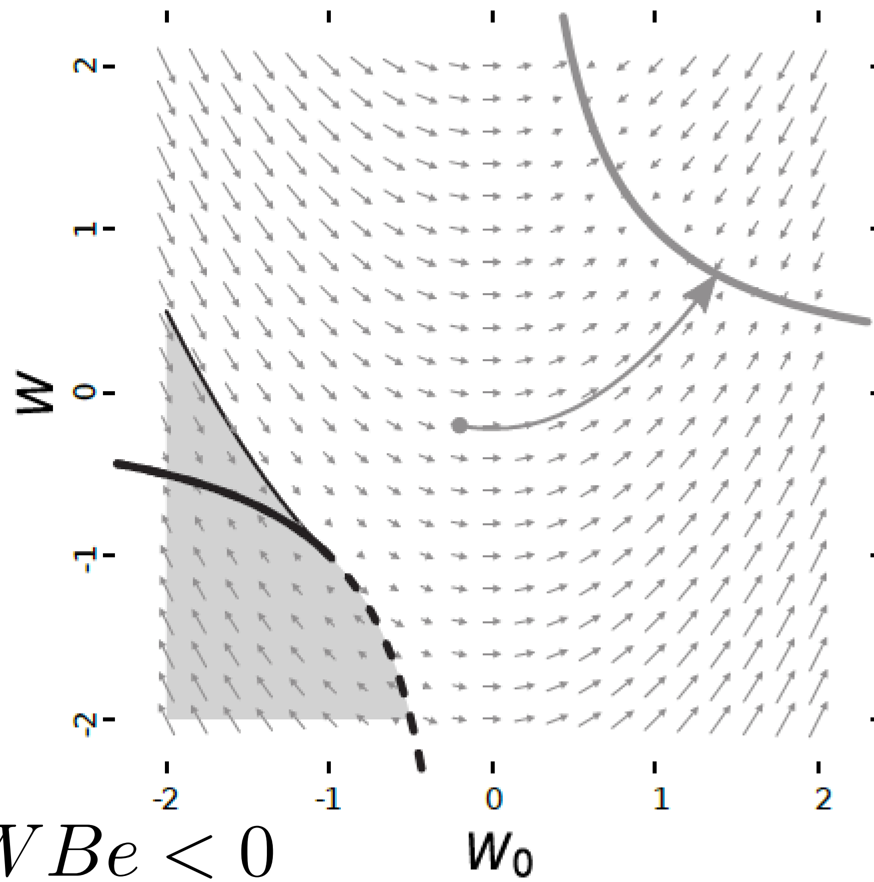
$$\Delta h_{BP}^T \Delta h_{FA} > 0$$

- All you need is to adjust W

Toy Example (1d)



- data: x=y
- Fix B=1



$eWBe > 0$

$W_0W = 1(\text{hyperbole})$

$eWBe < 0$

Simulations

- As good as back-propagation
- Decrease of $\Delta h_{FA} \measuredangle \Delta h_{BP}$

Analysis

Setting

$\tilde{y} = Tx$ (data)

$h = Ax$

$y = Wh$

$E = T - WA$ (error)

Learning

$$\langle\langle \dot{W} = EA^{T} \bar{A}^{T T^{T}} \rangle A^{T}$$

$$\langle \dot{A} = BE \eta BE xx^{T} \rangle$$

Theorem 1. *Given the learning dynamics and that the matrix* B *satisfies* $B^{+}B = I$

then $\lim_{t \to \infty} E = 0.$

Analysis

Proof #2: Gauss-Newton modification of backprop

Here we will show that replacing the transpose matrix, W^T, with the Moore-Penrose pseudoinverse matrix, W^+, in the backprop algorithm renders an update rule which approximates Gauss-Newton optimization. That is, the pseudoinverse of the forward matrix, W^+, not only satisfies the first condition from the main text, i.e. $\boldsymbol{e}^T W W^+ \boldsymbol{e} > 0$, it prescribes second order updates for the hidden units.

That's it!

Analysis

Proof #3: B acts like the pseudoinverse of W

Here we will prove that, under fairly restrictive conditions, feedback alignment prescribes hidden unit updates which are in the same direction as those prescriped by psuedobackprop, i.e. $\Delta \boldsymbol{h}_{\mathrm{FA}} \measuredangle \Delta \boldsymbol{h}_{\mathrm{PBP}} = 0$.