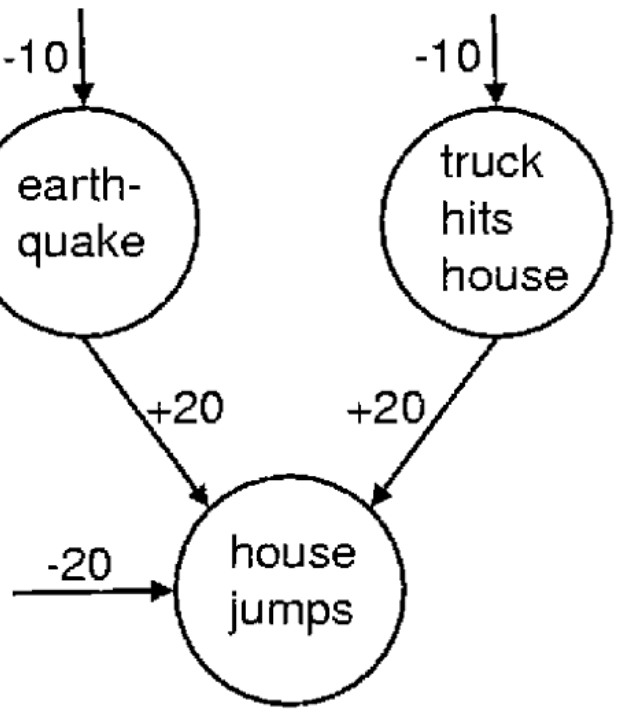
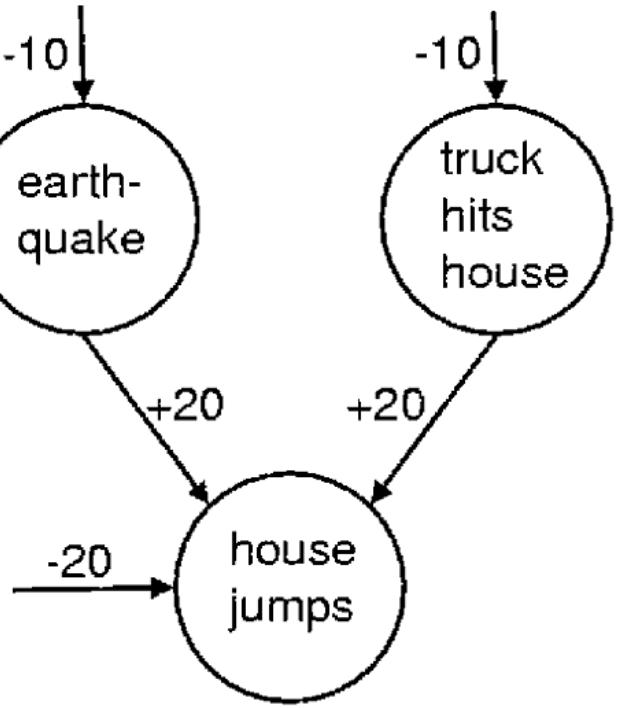
A Fast Learning Algorithm for Deep Belief Nets

Hinton, Osindero, Teh

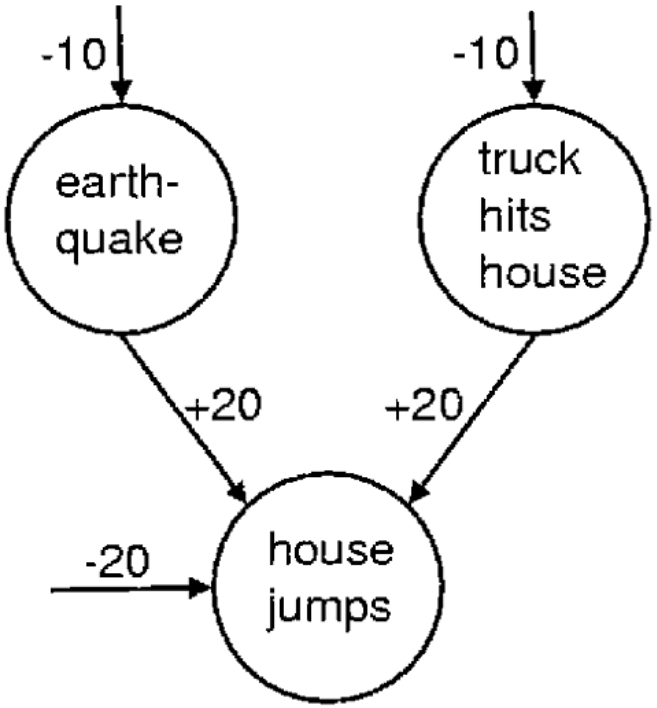
Conditional Learning is Hard



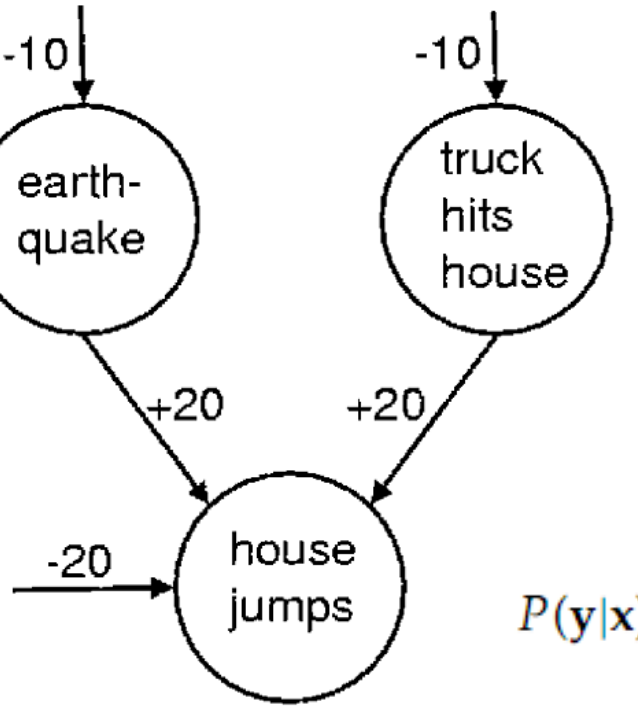
Conditional Learning is Hard



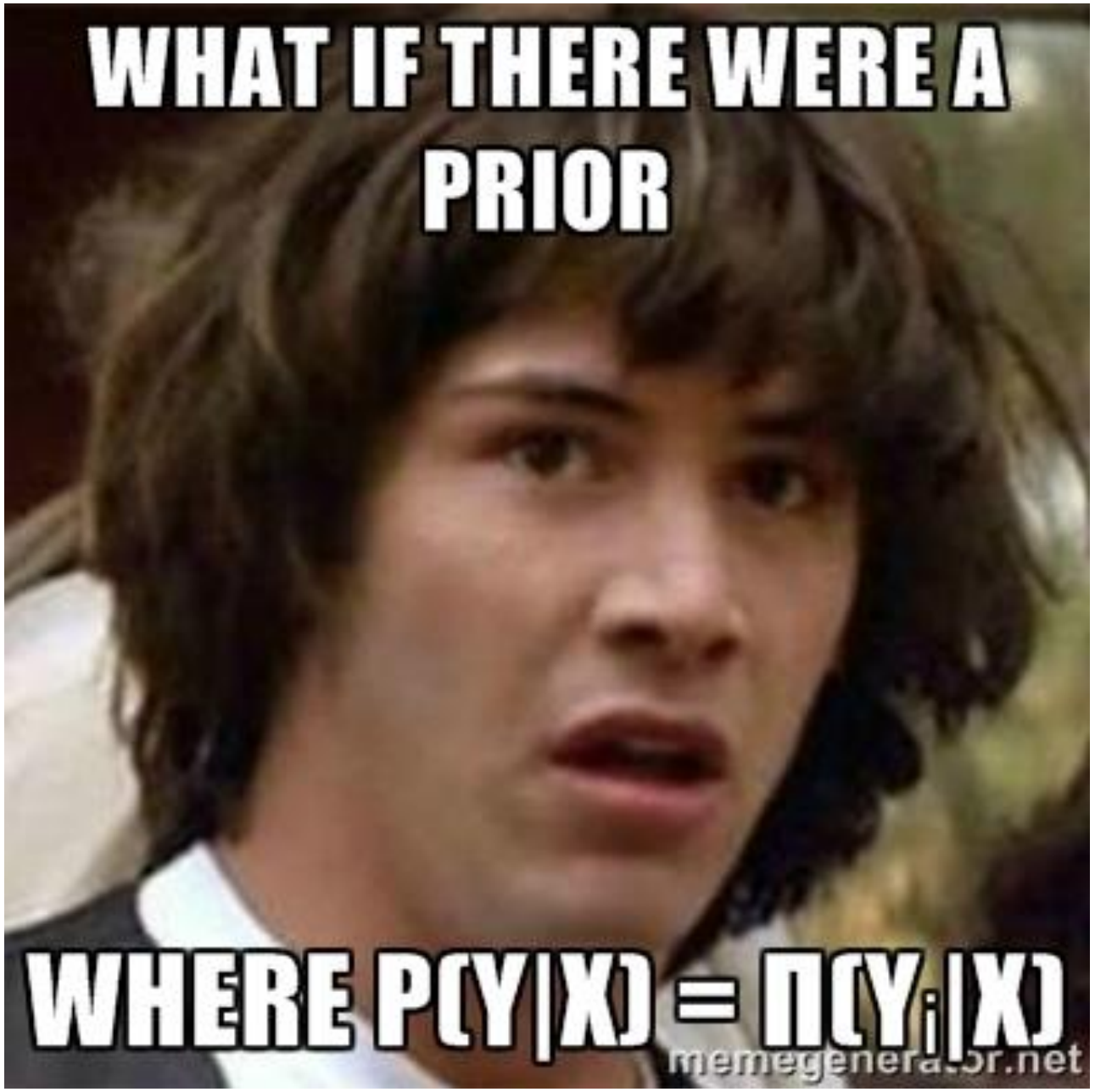
Conditional Learning is Hard



Conditional Learning is Hard



$$P(\mathbf{y}|\mathbf{x}) = \prod_j P(y_j|\mathbf{x})?$$

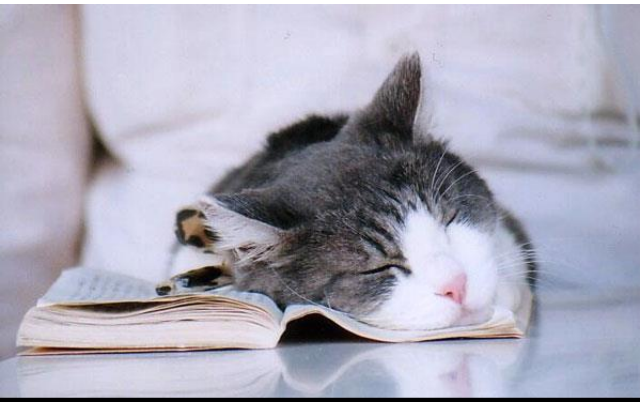

WHAT IF THERE WERE A PRIOR
WHERE P(Y|X) = Π(Yi|X)
memegenerator.net

Complementary Priors

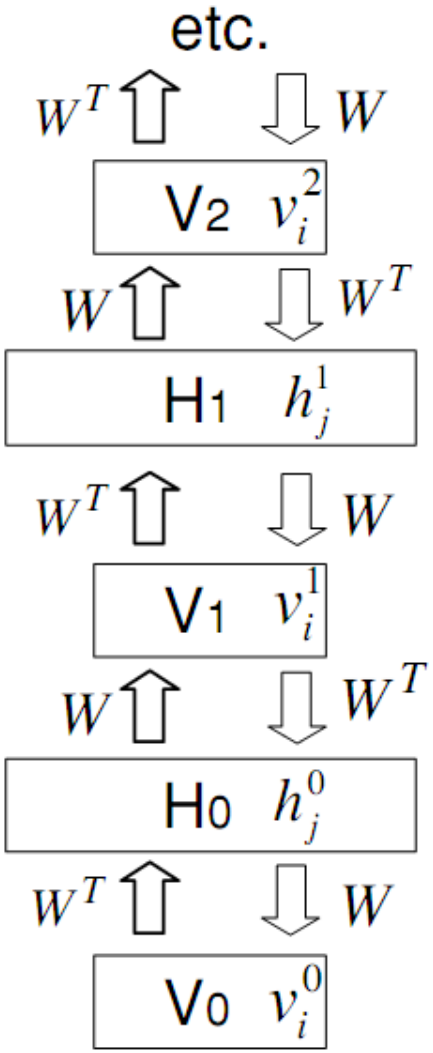
$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{\Omega(\mathbf{y})} \exp\left(\sum_i \Phi_j(\mathbf{x}, y_j) + \beta(\mathbf{x}) \right)$$

$$P(\mathbf{y}) = \frac{1}{C} \exp\left(\log \Omega(\mathbf{y}) + \sum_j \alpha_j(y_j) \right),$$

$$P(\mathbf{x}, \mathbf{y}) = \frac{1}{C} \exp\left(\sum_j \Phi_j(\mathbf{x}, y_j) + \beta(\mathbf{x}) + \sum_j \alpha_j(y_j) \right).$$



A specially structured deep network



Training our deep network

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}^{00}} = \langle h_j^0 (v_i^0 - \hat{v}_i^0) \rangle,$$

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}^{00}} = \langle h_j^0 (v_i^0 - v_i^1) \rangle.$$

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}} = \langle h_j^0 (v_i^0 - v_i^1) \rangle + \langle v_i^1 (h_j^0 - h_j^1) \rangle + \langle h_j^1 (v_i^1 - v_i^2) \rangle + \cdots$$

Training our deep network

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}} = \langle h_j^0 (v_i^0 - v_i^1) \rangle + \langle v_i^1 (h_j^0 - h_j^1) \rangle + \langle h_j^1 (v_i^1 - v_i^2) \rangle + \cdots$$

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}} = \langle v_i^0 h_j^0 \rangle - \langle v_i^\infty h_j^\infty \rangle.$$

Training our deep network

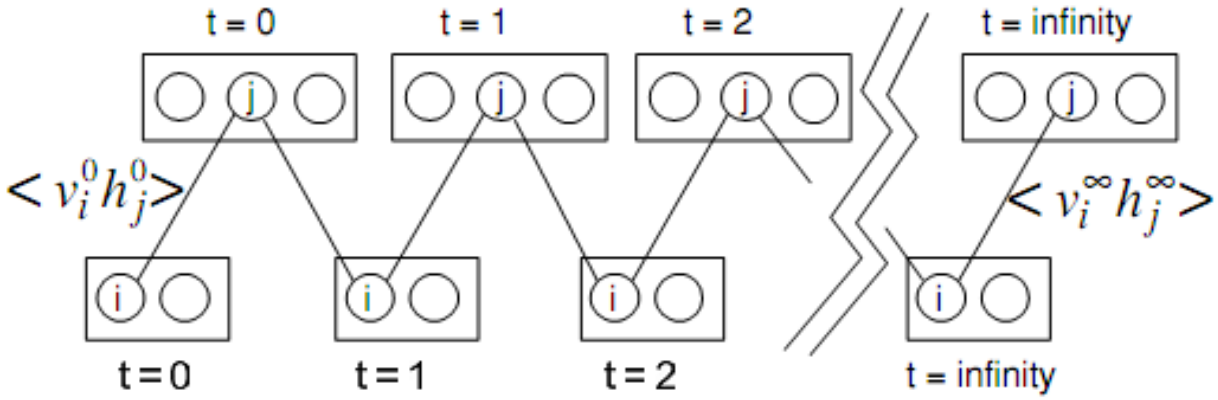
$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}} = \langle h_j^0 (v_i^0 - v_i^1) \rangle + \langle v_i^1 (h_j^0 - h_j^1) \rangle + \langle h_j^1 (v_i^1 - v_i^2) \rangle + \cdots$$

$$\frac{\partial \log p(\mathbf{v}^0)}{\partial w_{ij}} = \langle v_i^0 h_j^0 \rangle - \langle v_i^\infty h_j^\infty \rangle.$$

$$KL(P^0 \| P_\theta^\infty) - KL(P_\theta^n \| P_\theta^\infty).$$

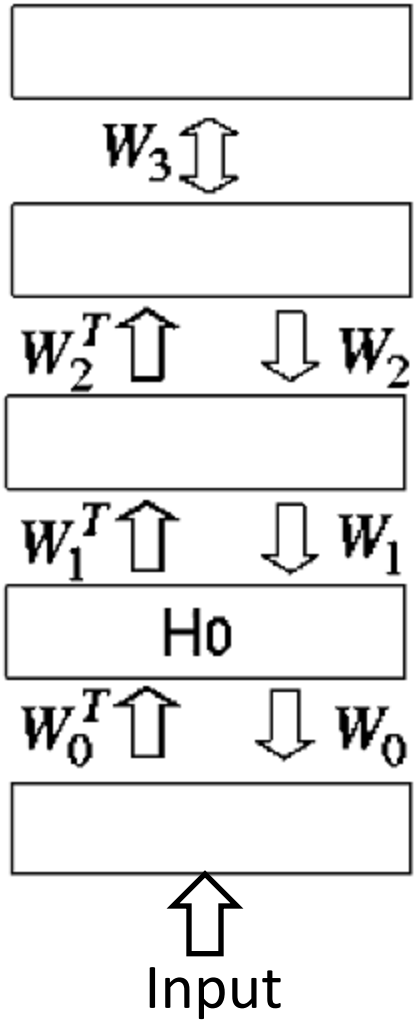
- This is the update for a restricted Boltzmann Machine

RBM training

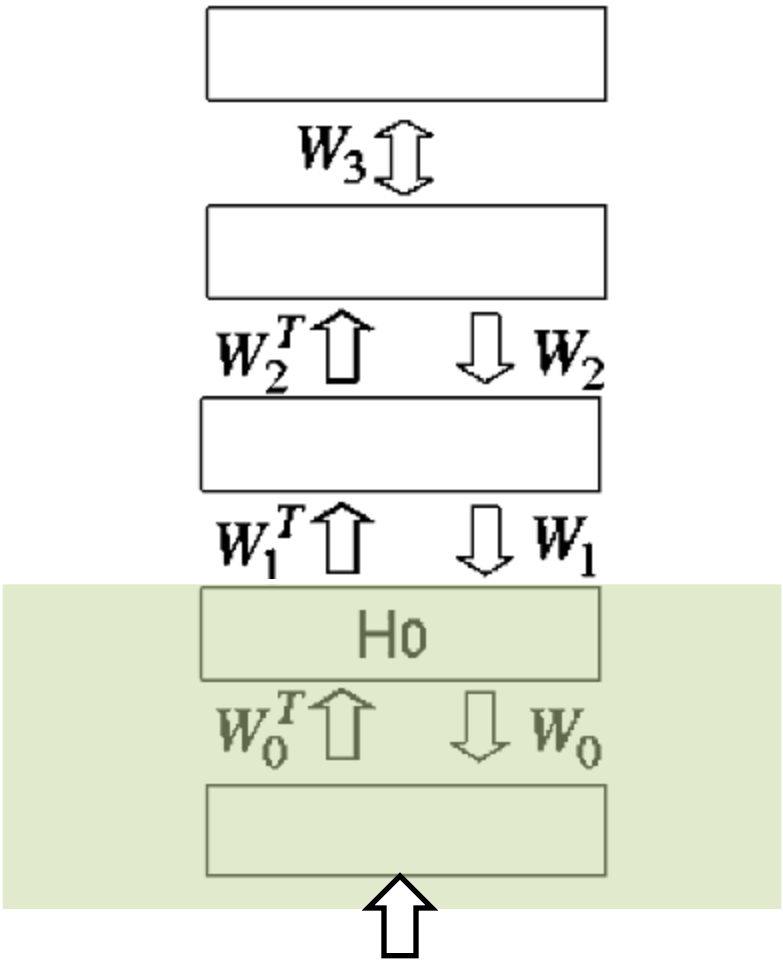


Let’s relax the assumptions

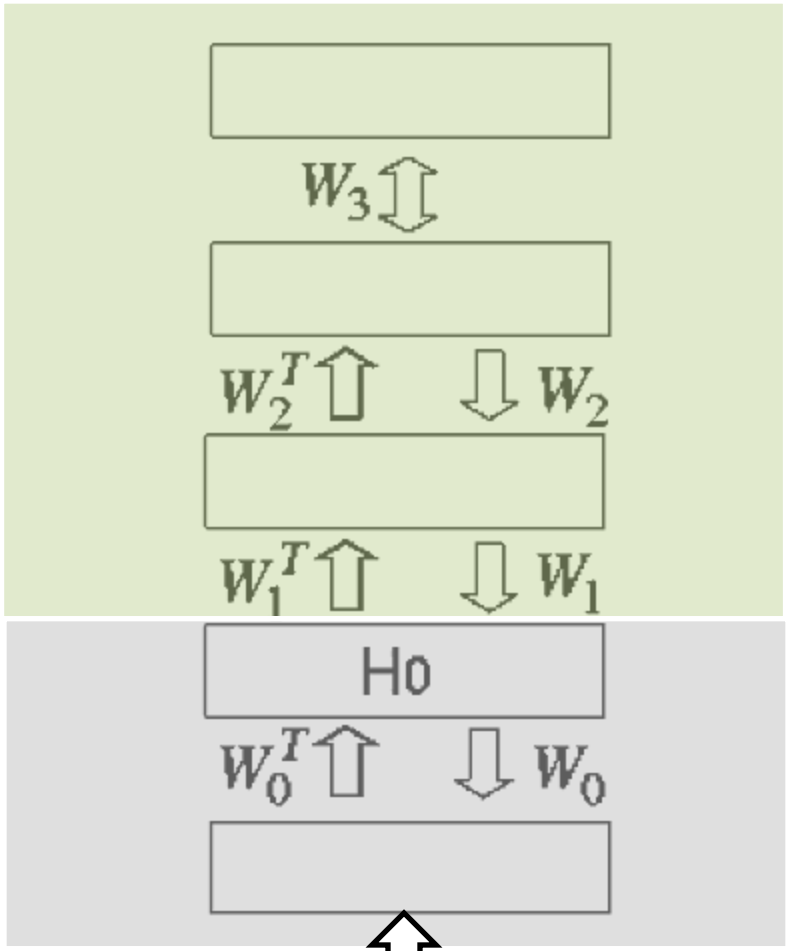
RBM



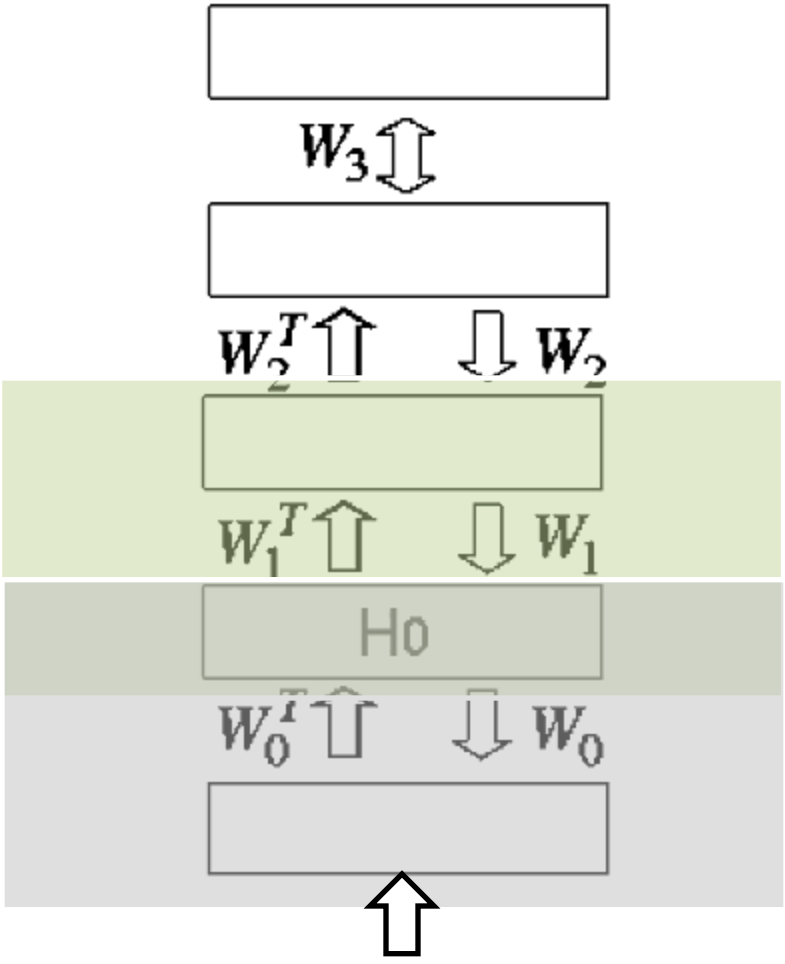
Greedy Training



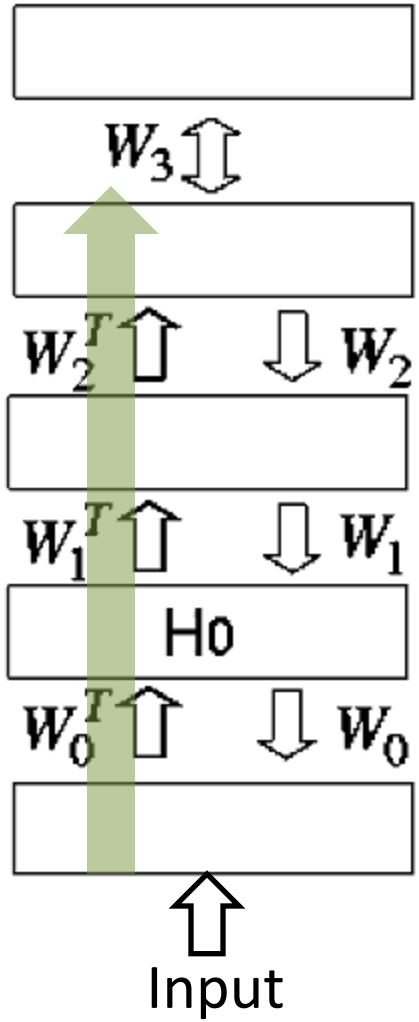
Greedy Training



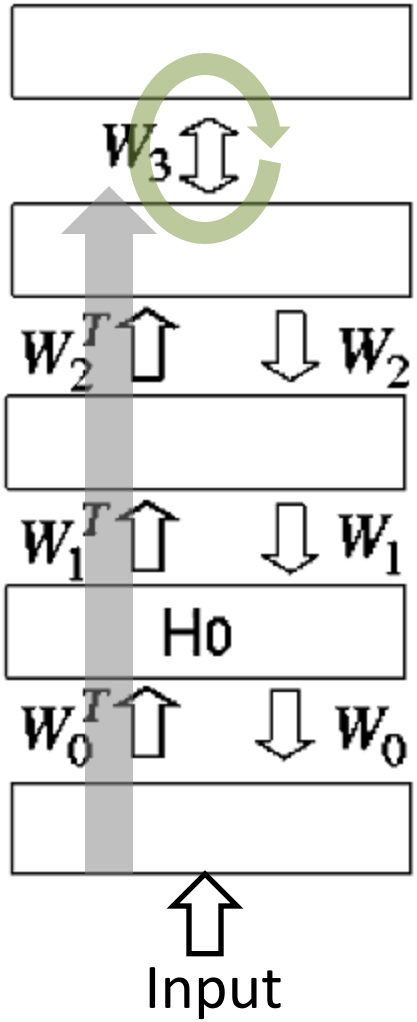
Greedy Training



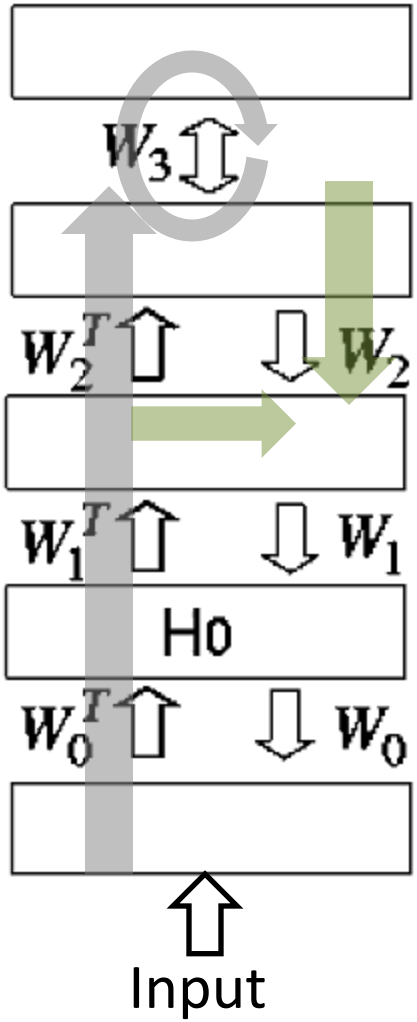
Fine tuning



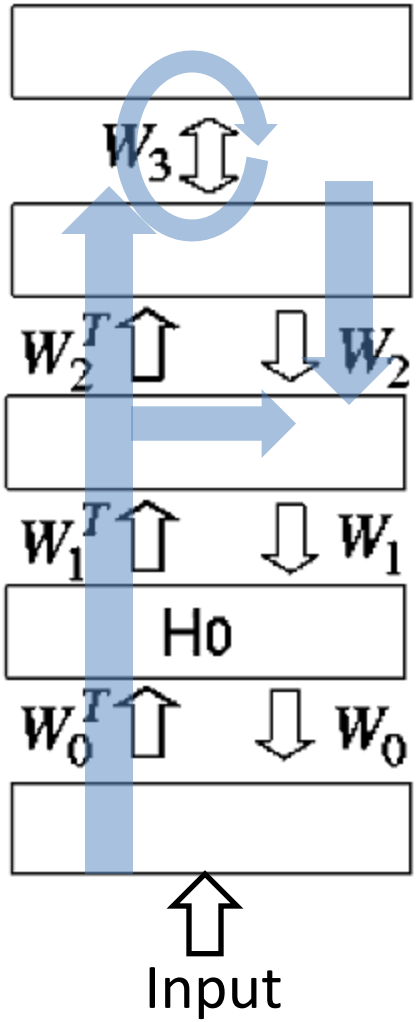
Fine tuning

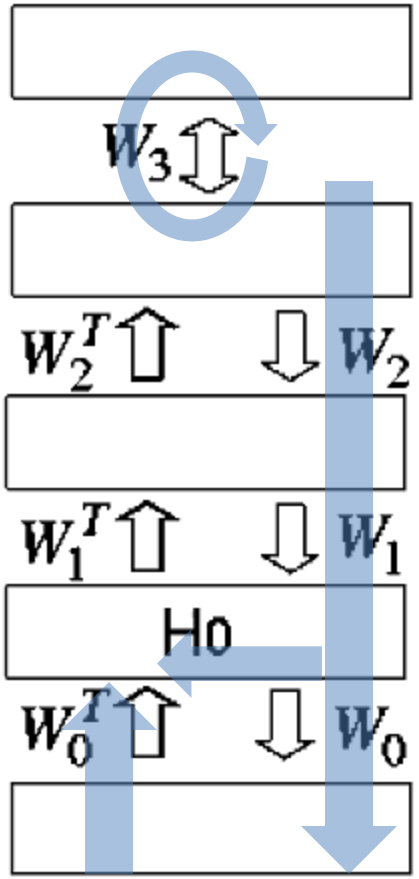


Fine tuning



Fine tuning





Yes, this actually works

All Done!