Structured and Deep Similarity Matching via Structured and Deep Hebbian Networks

Dina Obeid, Hugo Ramambason and Cengiz Pehlevan



Hebbian plasticity implements gradient-based learning

Overview

- We introduce structured and deep similarity matching cost functions
- We show how they can be optimized in a gradient-based manner by structured and deep neural networks with local learning rules.
- ©Credit assignment problem is solved elegantly by a factorization of the dual learning objective to synapse specific local objectives.
- 4 Simulations show that our networks learn meaningful features.

Introduction

- How can the brain's synapses, with access to only local information about the network, can do gradient-based optimization of an objective function?
- Researchers have been tackling this problem by searching for a biologically-plausible implementation of the backpropagation algorithm, but a fully plausible implementation is not yet available [3].
- We focus on networks already operating with biologically-plausible learning rules.
- We ask whether one can formulate network-wide learning cost functions for such networks and whether these networks achieve efficient "credit assignment" by performing gradient-based learning.

Regularized similarity matching and gradient-based learning in a Hebbian/anti-Hebbian network

Regularized similarity matching:

Inputs: $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{R}^K$, and outputs: $\mathbf{r}_1, \dots, \mathbf{r}_T \in \mathbb{R}^N$.

$$\min_{\mathbf{r}_{1},\dots,\mathbf{r}_{T}} \frac{1}{2T^{2}} \int_{t=1}^{T} \int_{t'=1}^{T} (\mathbf{x}_{t} \cdot \mathbf{x}_{t'} - \mathbf{r}_{t} \cdot \mathbf{r}_{t'})^{2} + \frac{2}{T} \int_{t=1}^{T} ||\mathbf{F}(\mathbf{r}_{t})||_{1},$$
s.t. $a \leq \mathbf{r}_{t} \leq b, \qquad t = 1,\dots,T.$ (1)

Here, the bounds and the regularization function act elementwise. **Dual problem:** Introduce new auxiliary variables $\mathbf{W} \in \mathbb{R}^{N \times K}$ and $\mathbf{L} \in$

$$-\frac{1}{T^{2}} \sum_{t} \sum_{t'} \mathbf{x}_{t'}^{\top} \mathbf{x}_{t'} \mathbf{r}_{t'}^{\top} \mathbf{r}_{t'} = \min_{\mathbf{W} \in \mathbb{R}^{N \times K}} -\frac{2}{T} \sum_{t} \mathbf{x}_{t}^{\top} \mathbf{W}^{\top} \mathbf{r}_{t} + \operatorname{Tr} \mathbf{W}^{\top} \mathbf{W},$$

$$\frac{1}{2T^{2}} \sum_{t} \sum_{t'} \left(\mathbf{r}_{t}^{\top} \mathbf{r}_{t'} \right)^{2} = \max_{\mathbf{L} \in \mathbb{R}^{N \times N}} \frac{1}{T} \sum_{t} \mathbf{r}_{t}^{\top} \mathbf{L} \mathbf{r}_{t} - \frac{1}{2} \operatorname{Tr} \mathbf{L}^{\top} \mathbf{L}.$$
(2)

A dual min-max formulation of similarity matching:

$$\min_{\mathbf{W} \in \mathbb{R}^{N \times K}} \max_{\mathbf{L} \in \mathbb{R}^{N \times N}} \frac{1}{T} \sum_{t=1}^{T} l_t(\mathbf{W}, \mathbf{L}, \mathbf{x}_t), \tag{3}$$

where

$$l_t := \operatorname{Tr} \mathbf{W}^{\top} \mathbf{W} - \frac{1}{2} \operatorname{Tr} \mathbf{L}^{\top} \mathbf{L} + \min_{\mathbf{r}_t} \left(-2\mathbf{r}_t^{\top} \mathbf{W} \mathbf{x}_t + \mathbf{r}_t^{\top} \mathbf{L} \mathbf{r}_t + 2 \| \mathbf{F}(\mathbf{r}_t) \|_1 \right).$$
(4)

Földiak's Hebbian/anti-Hebbian network [1]:

In the first step, the algorithm minimizes l_t with respect to \mathbf{r}_t by running the neural dynamics (7) until convergence. Minimization is achieved because the argument of min in (4), $E = -2\mathbf{r}_t^{\mathsf{T}}\mathbf{W}\mathbf{x}_t + \mathbf{r}_t^{\mathsf{T}}\mathbf{L}\mathbf{r}_t + 2\|\mathbf{F}(\mathbf{r}_t)\|_1$, is a Lyapunov function on the neural dynamics (7), if, within the bounds on the output, the regularizer is related to the neural activation function as:

$$F'(r) = u - r, \quad \text{where } r = f(u). \tag{5}$$

Explicit credit assignment:

$$\sum_{i=1}^{N} \sum_{j=1}^{K} \left(-2W_{ij}r_{t,i}^*x_{t,j} + W_{ij}^2\right) + \sum_{i=1}^{N} \sum_{j=1}^{N} \left(L_{ij}r_{t,i}^*r_{t,j}^* - \frac{1}{2}L_{ij}^2\right). \tag{6}$$

For each input $\mathbf{x} \in \mathbb{R}^K$, run until convergence:

$$\tau \frac{d\mathbf{u}(s)}{ds} = -\mathbf{u}(s) + \mathbf{W}\mathbf{x} - (\mathbf{L} - \mathbf{I})\mathbf{r}(s), \qquad \mathbf{r}(s) = \mathbf{f}(\mathbf{u}(s)). \tag{7}$$

Synaptic updates:

$$\Delta W_{ij} = \eta (r_i x_j - W_{ij}), \quad \Delta L_{ij} = \frac{\eta}{2} (r_i r_j - L_{ij}).$$
 (8)

Structured Similarity Matching

$$\min_{\substack{a \leq \mathbf{r}_{t} \leq b, \\ t=1, \dots, T}} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left[-\sum_{i,j} x_{t,i} x_{t',i} r_{t,j} r_{t',j} c_{ij}^{W} + \frac{1}{2} \sum_{i,j} r_{t,i} r_{t',i} r_{t,j} r_{t',j} c_{ij}^{L} \right] + \frac{2}{T} \sum_{t=1}^{T} \left\| \mathbf{F}(\mathbf{r}_{t}) \right\|_{1}.$$
(9)

Through the choice of c_{ij}^W and c_{ij}^L , one can design many topologies for the interactions between inputs and outputs, and outputs themselves.

Structured and Deep Similarity Matching

Deep Similarity Matching:

For notational convenience, we set $\mathbf{r}_t^{(0)} := \mathbf{x}_t$ and $N^{(0)} := K$, and define deep similarity matching with P layers as:

$$\min_{\substack{a \leq \mathbf{r}_{t}^{(p)} \leq b \\ t=1,\dots,P}} \sum_{p=1}^{P} \frac{\gamma^{p-P}}{2T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left(\mathbf{r}_{t}^{(p-1)} \cdot \mathbf{r}_{t'}^{(p-1)} - \mathbf{r}_{t}^{(p)} \cdot \mathbf{r}_{t'}^{(p)} \right)^{2} + \sum_{p=1}^{P} \frac{2\gamma^{p-P}}{T} \sum_{t=1}^{T} \left\| \mathbf{F} \left(\mathbf{r}_{t}^{(p)} \right) \right\|_{1}, \tag{10}$$

where $\gamma \geq 0$ is a parameter and $\mathbf{r}_t^{(p)} \in \mathbb{R}^{N^{(p)}}$.

Structured and deep similarity matching:

Structured and deep similarity matching:

$$\min_{\substack{a \leq \mathbf{r}_{t}^{(p)} \leq b \\ t=1,...,P}} \frac{P}{T^{2}} \frac{\gamma^{p-P}}{T^{2}} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left(-\frac{\sum_{i=1}^{N(p-1)} N_{t,i}^{(p)}}{\sum_{i=1}^{S} \sum_{j=1}^{T}} r_{t,i}^{(p-1)} r_{t',i}^{(p)} r_{t',j}^{(p)} c_{ij}^{W,(p)} \right) + \frac{2\gamma^{p-P}}{T} \sum_{t=1}^{T} \left\| \mathbf{F} \left(\mathbf{r}_{t}^{(p)} \right) \right\|_{1}^{N(p)} N_{t,i}^{(p)} r_{t,i}^{(p)} r_{t,i}^{(p)} r_{t,j}^{(p)} r_{t,j}^{(p)} c_{ij}^{L,(p)} \right) + \frac{P}{p-1} \frac{2\gamma^{p-P}}{T} \sum_{t=1}^{T} \left\| \mathbf{F} \left(\mathbf{r}_{t}^{(p)} \right) \right\|_{1}^{N(p)} N_{t,i}^{(p)} r_{t,i}^{(p)} r_{t,i}^{(p)} r_{t,j}^{(p)} r_{t,j}^{$$

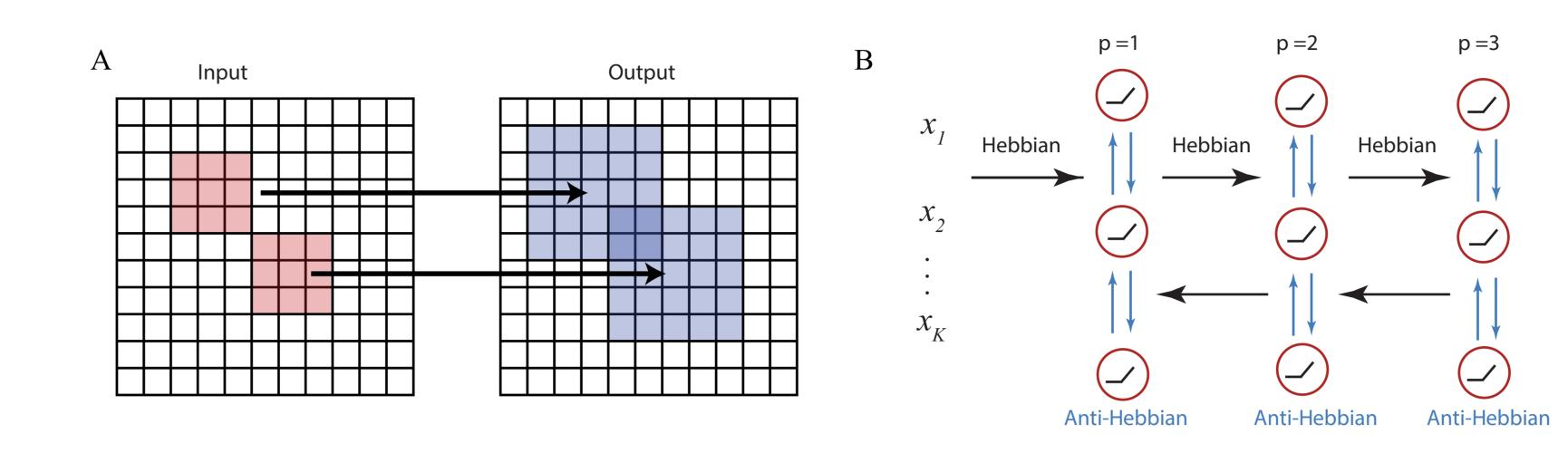


Figure 1: A) Locally connected similarity matching and structured Hebbian/anti-Hebbian network. B) Deep similarity matching and deep Hebbian/anti-Hebbian network. Arrows illustrate synaptic connections. One can introduce structure for all components of the con-

Structured and Deep Hebbian/anti-Hebbian Networks

We introduce dual variables $W_{ij}^{(p)}$ and $L_{ij}^{(p)}$ for interactions with positive structure constants, define variables

$$\bar{W}_{ij}^{(p)} = \begin{cases} W_{ij}^{(p)}, c_{ij}^{W,(p)} \neq 0 \\ 0, c_{ij}^{W,(p)} = 0 \end{cases}, \qquad \bar{L}_{ij}^{(p)} = \begin{cases} L_{ij}^{(p)}, c_{ij}^{L,(p)} \neq 0 \\ 0, c_{ij}^{L,(p)} = 0 \end{cases}, \tag{12}$$

for notational convenience, and rewrite (11) as

$$\min_{\bar{\mathbf{W}}^{(1)}, \dots, \bar{\mathbf{W}}^{(p)}} \max_{\bar{\mathbf{L}}^{(1)}, \dots, \bar{\mathbf{L}}^{(p)}} \frac{1}{T} t_{=1}^{T} l_t(\bar{\mathbf{W}}^{(1)}, \dots, \bar{\mathbf{W}}^{(p)}, \bar{\mathbf{L}}^{(1)}, \dots, \bar{\mathbf{L}}^{(p)}, \mathbf{r}_t^{(0)}), \qquad (13)$$

$$t := \sum_{p=1}^{P} \sum_{\substack{i,j \\ c_{ij}^{W,(p)} \neq 0}} \frac{\gamma^{p-P}}{c_{ij}^{W,(p)}} W_{ij}^{(p)^{2}} - \sum_{p=1}^{P} \sum_{\substack{i,j \\ c_{ij}^{L,(p)} \neq 0}} \frac{\gamma^{p-P}}{2(1+\gamma(1-\delta_{pP})) c_{ij}^{L,(p)}} L_{ij}^{(p)^{2}} + \min_{\substack{a \leq \mathbf{r}_{t}^{(p)} \leq b \\ p=1,...,P}} \sum_{p=1}^{P} \gamma^{p-P} \left[-2\mathbf{r}_{t}^{(p)^{\top}} \mathbf{\overline{W}}^{(p)} \mathbf{r}_{t}^{(p-1)} + \mathbf{r}_{t}^{(p)^{\top}} \mathbf{\overline{L}}^{(p)} \mathbf{r}_{t}^{(p)} + 2 \left\| \mathbf{F} \left(\mathbf{r}_{t}^{(p)} \right) \right\|_{1} \right],$$

$$(14)$$

Neural dynamics

Run the following neural network dynamics until convergence,

$$\tau \frac{d\mathbf{u}^{(p)}}{ds} = -\mathbf{u}^{(p)} + \bar{\mathbf{W}}^{(p)}\mathbf{r}^{(p-1)} - (\bar{\mathbf{L}}^{(p)} - \mathbf{I})\mathbf{r}^{(p)} + (1 - \delta_{pP})\gamma \bar{\mathbf{W}}^{(p+1)\top}\mathbf{r}^{(p+1)},$$

$$\mathbf{r}^{(p)} = \mathbf{f}(\mathbf{u}^{(p)}), \quad p = 1, \dots, P.$$
(15)

Gradient-based learning and local learning rules

$$l_{t} = \sum_{p=1}^{P} \sum_{\substack{i,j \ c_{ij}^{W,(p)} \neq 0}} \left[-2W_{ij}^{(p)} r_{j}^{(p)} r_{i}^{(p-1)} + \frac{\gamma^{p-P}}{c_{ij}^{W,(p)}} W_{ij}^{(p)^{2}} \right]$$

$$- \sum_{p=1}^{P} \sum_{\substack{i,j \ c_{ij}^{L,(p)} \neq 0}} \left[-L_{ij}^{(p)} r_{j}^{(p)*} r_{i}^{(p)*} + \frac{\gamma^{p-P}}{2(1+\gamma(1-\delta_{pP})) c_{ij}^{L,(p)}} L_{ij}^{(p)^{2}} \right].$$
 (16)

Local learning rules are derived from the above equation by taking deriva-

$$\Delta W_{ij}^{(p)} = \eta \gamma^{p-P} \left[r_j^{(p)} r_i^{(p-1)} - \frac{W_{ij}^{(p)}}{c_{ij}^{W,(p)}} \right],$$

$$\Delta L_{ij}^{(p)} = \frac{\eta}{2} \gamma^{p-P} \left[r_j^{(p)} r_i^{(p)} - \frac{L_{ij}^{(p)}}{(1 + \gamma(1 - \delta_{pP})) c_{ij}^{L,(p)}} \right]. \tag{17}$$

Simulations

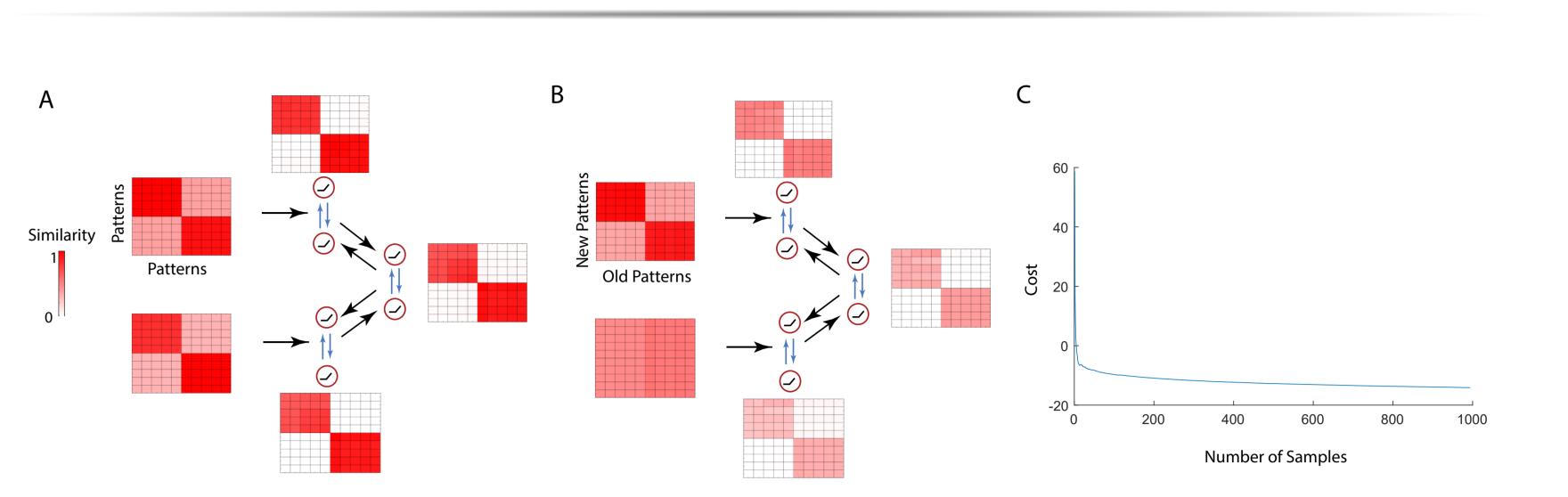


Figure 2: A two-layer Hebbian/anti-Hebbian network with feedback. For each subnetwok, representational similarity matrices are shown. Similarities are calculated by taking pairwise dot products of patterns and normalizing the largest dot product to $1.\,A)$ Network simulated with patterns from a set generated from the same distribution as the training set. B) Network simulated with patterns to the bottom first layer generated from a different distribution. C) Structured and deep similarity matching cost decreases over training.

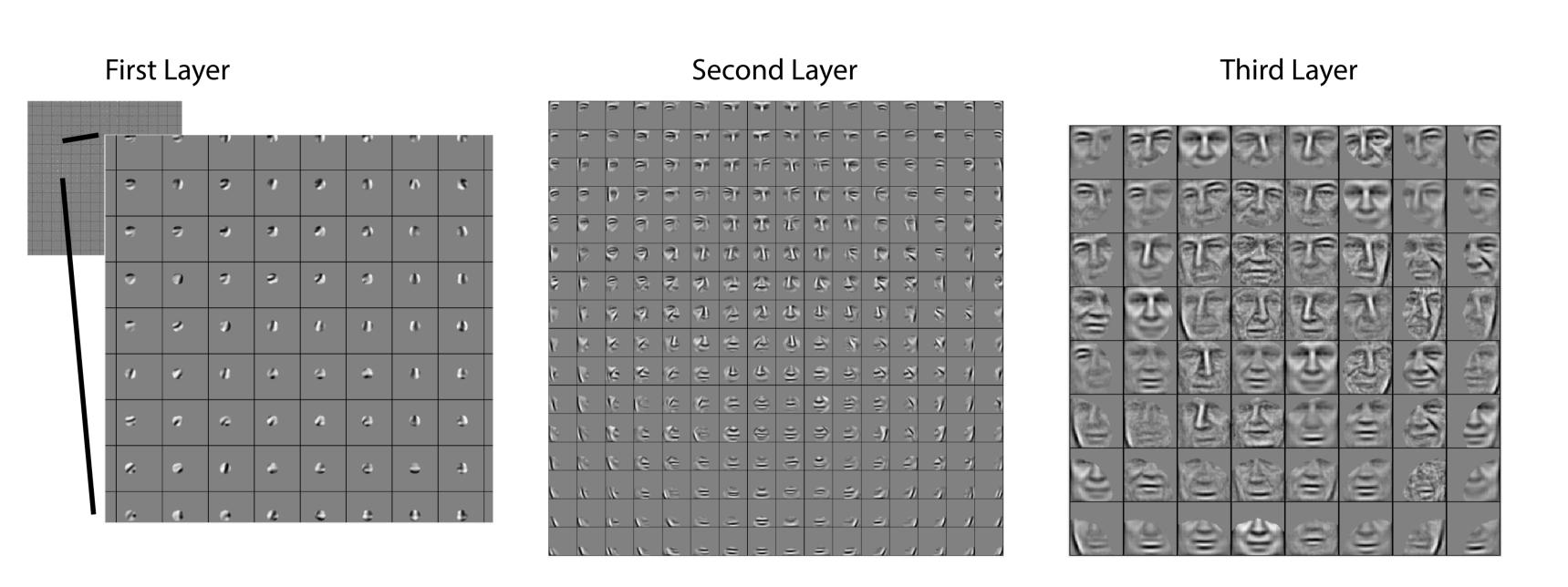


Figure 3: Features learned by a 3-layer, locally connected Hebbian/anti-Hebbian neural network on the labeled faces in the wild dataset [2]. Features are calculated by reverse correlation on the dataset, and masking these features to keep only the portions of the dataset which elicits a response in the neuron.

Table 1: Classification on MNIST data set: we show how the test error decreases as the number of neurons per site (NPS) increases.

References

[1] P. Földiak.

Forming sparse representations by local anti-hebbian learning. $Biological\ cybernetics,\ 64(2):165-170,\ 1990.$

[2] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua. Labeled faces in the wild: A survey.

In Advances in face detection and facial image analysis, pages 189–248. Springer,

[3] J. C. Whittington and R. Bogacz.

Theories of error back-propagation in the brain.

Trends in cognitive sciences, 2019.

Acknowledgements