Plasticity rules for learning sequential inputs under energetic constraints

Dmytro Grytskyy[1], **Renaud Jolivet**[1,2]

[1]Département de physique nucléaire et corpusculaire (DPNC), University of Geneva, Switzerland [2]CERN, Geneva, Switzerland

Dmytro.Grytskyy@etu.unige.ch

1. Objective and Summary

Information measures are often used to obtain the efficiacy of neural networks, and learning rules can be derived as optimization procedures of some information measure [4]. On the other side, it is bounded by amount of avaliable resources [5]. Here we study learning rules balancing information inference and spent energy.We first consider time local inputs and then develop a generalization for time sequences of inputs. We:

- set goal function equal to the mutual information between external input and network's output
- derive learning rules as a gradient descent optimizing this function
- add term proportional to spent energy to the goal function
- take synaptic noise and resulting unreliability of incoming signals into account

2. Results

- derived learning rule consists of in time and space local terms and a nonlocal value common for all neurons
- also sequences can be learned
- more often cases occupy lower energy orbits
- too inessential items are dropped off if noise taken into account

3. Model&Methods

- networks of Hawkes neurons with reset after spiking
- considered dynamics is $\partial_t u_i = [W_{ij}y_j + V_{ik}x_k] - \frac{1}{\tau}u_i$, $p(y=1) = f(u)$
- $u_i = W_{ij}s_j^y + V_{ik}s_k^x$ with $s_l^z = \int_0^{t-t_i^{sp}} z_l(t-\tau)h(\tau)d\tau$ with stepwise h for inputs without time dependencies and $h = \tau^{-1}\exp(-t/\tau)$ for inputs developing in time
- mutual information:
  $M = \sum_x p(x)\sum_y p(y|x)\ln p(y|x) - \sum_y \sum p(x)p(y|x)\ln\{\sum p(x)p(y|x)\}$
- energy term: $E = \underbrace{\sum_k y_k}_{E_{sp}} + c_{psp}\underbrace{\sum_{i,j,k}(W_{ij}y_j + V_{ik}x_k)}_{E_{psp}}$
- minimization of $F = M - \gamma E_{sp} - \underbrace{\gamma c_{psp}}_{\gamma_{psp}} E_{psp}$ instead of just M provides additional terms to learning rule: $\partial_{w_{ij}}E = \int D\epsilon \partial_{w_{ij}} p(y[t]) \int \sum_k y_k dt + c_{psp}p(y_j = 1)$

4. Learning as optimization procedure

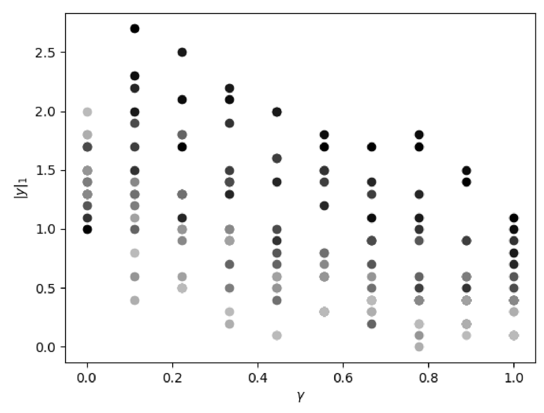
- Learning rule for $\partial_t w$ ($w = \{V, W\}$): optimize M as a gradient descent $\partial_t w = -\lambda\partial_{w_{ij}}F$
- For in time disjoint inputs:
- Using $\ln p(y|x) = \sum \ln p_{i,t}$ with $p_i = f(u_i)$ for $y_i = 1$ and $p_i = 1 - f(u_i)$ for $y_i = 0$: $\partial_{w_{ij}}M = \sum_x\sum_y p(x)\{p(y|x)/p_{i|x}x_j(2y_i - 1)f^/(u_i) * (\ln p(y|x) - \ln p(y) - \gamma E_{sp}) - \gamma_{psp}\}$
- $\Delta w_{ij} = x_j[f^/(u_i)/p_i(2y_i - 1) * \{\ln p_{i,t} - \ln p(y) - \gamma E_{sp}\} - \gamma_{psp}]$ (approximating $p(y)$ with help of neurons activity covariances)
- For time sequences of inputs:
- Using $\ln p(y[t]|x[t]) = \sum \ln p_{k,t}$:

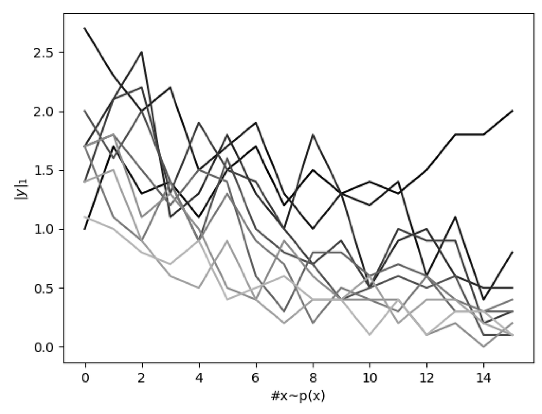
$\partial_{w_{ij}}M = \sum_x\sum_y p(x)p(y|x)[\{\sum_t(2y_{i,t} - 1)s_{i,t}^{y_j}f^/(u_i)/p_{i,t}\} * \{\sum \ln p_{k,t} - \ln p(y) - \gamma E_t\} - \gamma_{psp}\sum y_{j,t}]$ (with $\ln p(y) \sim \sum \ln p(y_t|y_{t-1..t_{sp_i}})$ for every pair of $\{x[t], y[t]\}$ of length T.

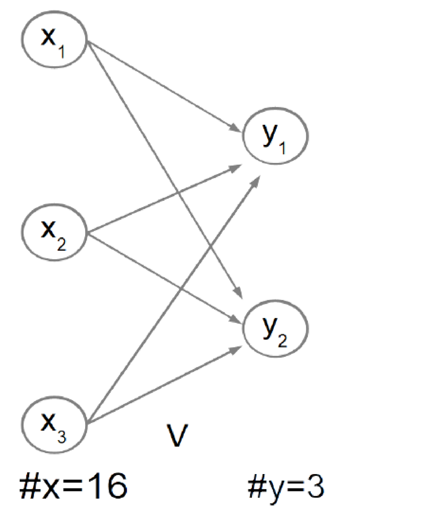
- Summing over many overlapping pairs : $\delta w_{ij} = \tilde{A}_{ij}B - \gamma_{psp}r_j$ with $\tilde{A}_{ij} = \int \Delta(t-\tau)A_{ij}(\tau)d\tau$, $\Delta(\tau) = |\tau|/T\theta(T - |\tau|)$, $A_{ij} = (2y_{i,t} - 1)s_{i,t}^{y_j}f^/(u_{i,t})/p_{i,t}$, $B = \ln p_{i,t} - \ln p(y_t|y_{t-1..t_{sp_i}}) - \gamma E_t$.

5. Interpretation of learning rules

- Sliding threshold for learning ~ BCM rules [7]
- More often x's occupy lower energy levels (analog to Huffman's coding [6]): $\Delta w \sim \ln p(y|x) - \ln p(y) - \gamma E$.
- Coding of coming x can rely on previous y's
  - For stable repeating sequences first spiking neuron inhibits further redundant spikes
- Global term (B) can be mediated by chemical communication [8]
- $y[t]$ not influenced by particular $y_{i,t}$ do not contribute to $\partial_t w_{ij}$: learning is causal
- Notions for external input are distinguished not only by time local input pattern, but also by following future
- If feedback from y to x added, learning supports neurons mostly influencing future x. It induces exploratory behavior.

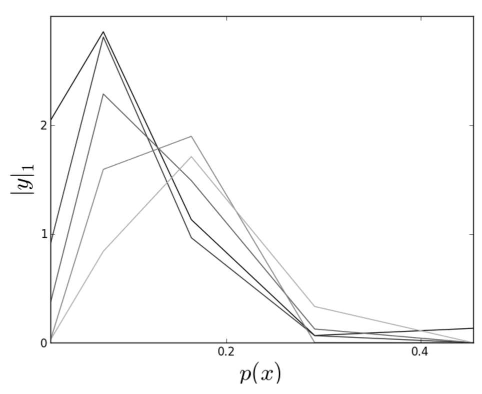


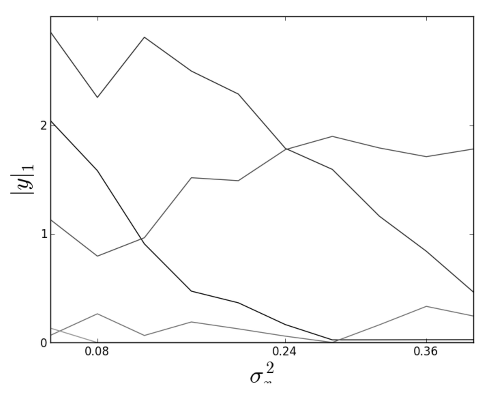




6. Noisy inputs

- For inputs $\tilde{x}_i = x_i + \xi_i$ with Gaussian noise ξ_i with variance σ_{nz}^2, $p(y_i = 1|x) = G(I_i, \sigma_{I_i}^2) = \frac{1}{\sqrt{2\pi}\sigma_I}\int f(I+\xi)\exp(-\frac{\xi^2}{2\sigma_I^2})d\xi$ with $\sigma_{I_i}^2 = \sum_i V_{ij}^2\sigma_{nz}^2$
- $\partial_{V_{ij}}M = \sum p_x\{x_j\partial_I G_i + \partial_{V_{ij}}\sigma_{I_i}\partial_\sigma G_i\}\left\{\ln\frac{G_i}{1-G_i} - \ln\frac{p(y_i=1)}{1-p(y_i=1)}\right\}$ with a new term cause of $\partial_{V_{ij}}\sigma_{I_i} = \frac{V_{ij}\sigma_j^2}{\sigma_I}$ keeping σ_I^2 and $V's$ away from big value
- With $\tilde{x}$, not x: in the leading order of σ_{nz}: $\partial_{V_{ij}}M \sim p(\tilde{x}, y)\{\tilde{x}_j\frac{\partial_I\tilde{p}(y_i|\tilde{x})}{\tilde{p}(y_i|\tilde{x})} - V_{ij}\sigma_j^2/\sigma_I^2\}\ln\frac{\tilde{p}(y|\tilde{x})}{p(y)}$ with the noise-induced term ~$V_{ij}\sigma_j^2/\sigma_I^2$.
- This term prevents learning rare inputs not distinguishable from noise. So, $|y|_1$ is strongest for moderate $p(x)$, rare, but still clear recognizable.





7. Outlook: further planned studies

- explicit separation of neurons in excitatory and inhibitory and sparseness effects
- oscillations as self-organized way of saving energy by population coding (similar to [2]), especially by storing information for unknown time
- subpopulations coding more abstract notions being energetically to expencive to code in connections between neurons coding lower abstraction level
- relation of self-organized criticallity, inference vs energy and excitaiton vs inhibition

Acknowledgments

This work is supported by the Swiss National Science Foundation (31003A_170079) and by the Australian Research Council (DP180101494).

.

References

[1] J. Brea,W. Senn, J.-P Pfister. (2011) Sequence learning with hidden units in spiking neural networks.NIPS 2011:1422-1430.
[2] M. Chalk, B. Gutkin, S. Deneve (2016) Neural oscillations as a signature of efficient coding in the presence of synaptic delays. eLife 5:e13824.
[3] G. Chechik (2003) Spike-Timing Dependent Plasticity and Relevant Mutual Information Maximization. Neural Computation, 15(7):1481-1510.
[4] Echeveste, Gros. (2014) Generating functionals for computational intelligence: The fisher information as an objective function for self-limiting hebbian learning rules. Front in Robotics and AI, 1:1.
[5] J. Harris, R. Jolivet, D. Attwell (2012) Synaptic Energy Use and Supply. Neuron 75:(5)762-777.
[6] D. Huffman (1952) A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE. 40 (9): 1098–1101.
[7] N. Intrator, L. Cooper (1992) Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. Neural Networks 5:3-17.
[8] S. Vincent (2010) Nitric oxide neurons and neurotransmission, Progress in Neurobiology 90(2):246-255.

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

CERN