BAYESIAN WEIGHT UPDATES STABILIZE AND IMPROVE LOCAL LEARNING IN A RECURRENT NEURAL NETWORK

Jorge A. Menéndez[1,2] & Peter E. Latham[1]

[1] Gatsby Computational Neuroscience Unit, [2] Centre for Computation, Mathematics and Physics in the Life Sciences and Experimental Biology, University College London

The Problem: local learning in the brain

How can a synapse know how to set its strength without knowing the others'?

solution: given the information locally available, estimate the strengths of the others and use this to estimate your own

➡ **Hypothesis:** synapses *optimally* integrate estimates of the others' weights to infer their own

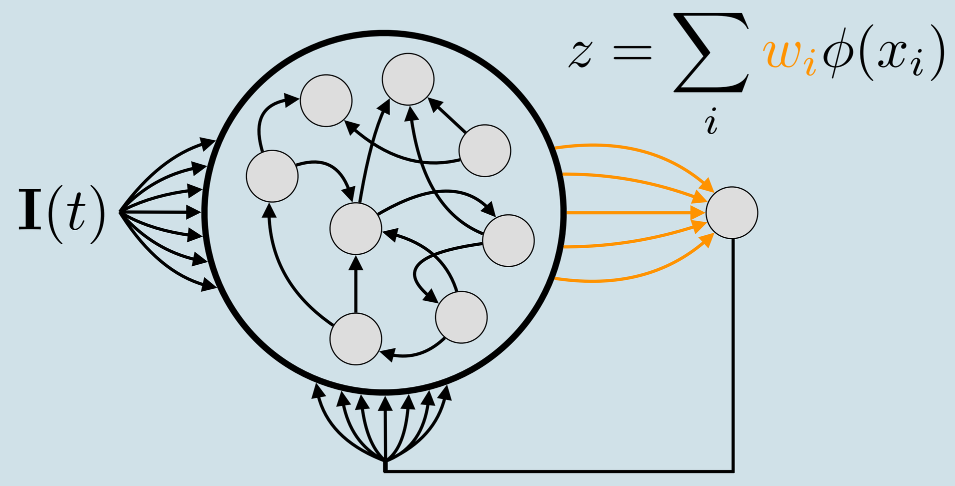
$$w_i(t+1) = \arg\max_w \log P(w_i^* = w | \underbrace{\mathcal{D}_i(t)}_{\text{locally available data}})$$

$$\approx w_i(t) + \underbrace{\sigma_i^2(t)}\,\mathcal{L}'_{i,t}(w_i(t))$$

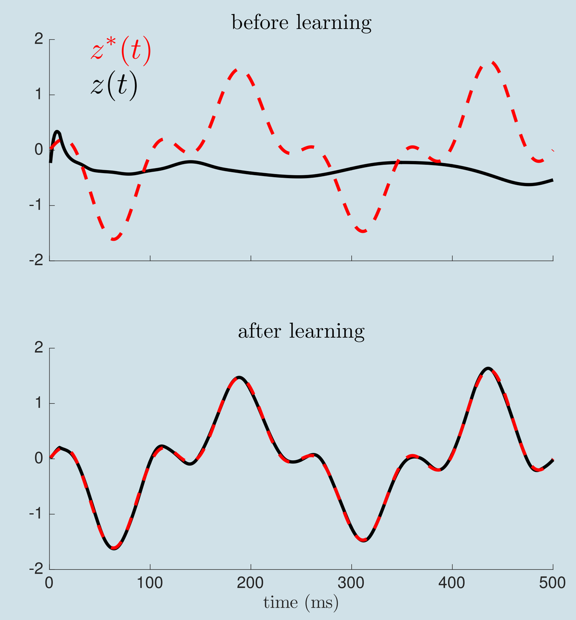
learning rate ∝ uncertainty

Q. How can this improve learning in a recurrent network?

Setup: reservoir computing



$$\tau\dot{\mathbf{x}} = -\mathbf{x} + \mathbf{J}\phi(\mathbf{x}) + \mathbf{u}z + \mathbf{I}$$



- every synapse has access to global supervisory error signal

$$\mathcal{D}_i(t) = \{\phi(x_i(t')), w_i(t'), \delta(t')\}_{t'=1}^t, \quad \delta(t) = z^*(t) - z(t)$$

- synapse i's model of other synapses:

$$\forall j \neq i, \quad \mathbb{E}[w_j^* - w_j] = 0, \quad \mathbb{V}\mathrm{ar}[w_j^* - w_j] = \sigma_\ell^2/N$$

- this implies Gaussian likelihood:

$$\delta(t) = (w_i^* - w_i)\phi(x_i) + \underbrace{\sum_{j\neq i}(w_j^* - w_j)\phi(x_j)}_{\to\mathcal{N}(0,\sigma_\ell^2)}$$

$$P(w_i^* = w|\mathcal{D}_i(t)) \propto P(\delta(t)|w_i^* = w, w_i(t), \phi(x_i(t))) \times P(w_i^* = w|\mathcal{D}_i(t-1))$$

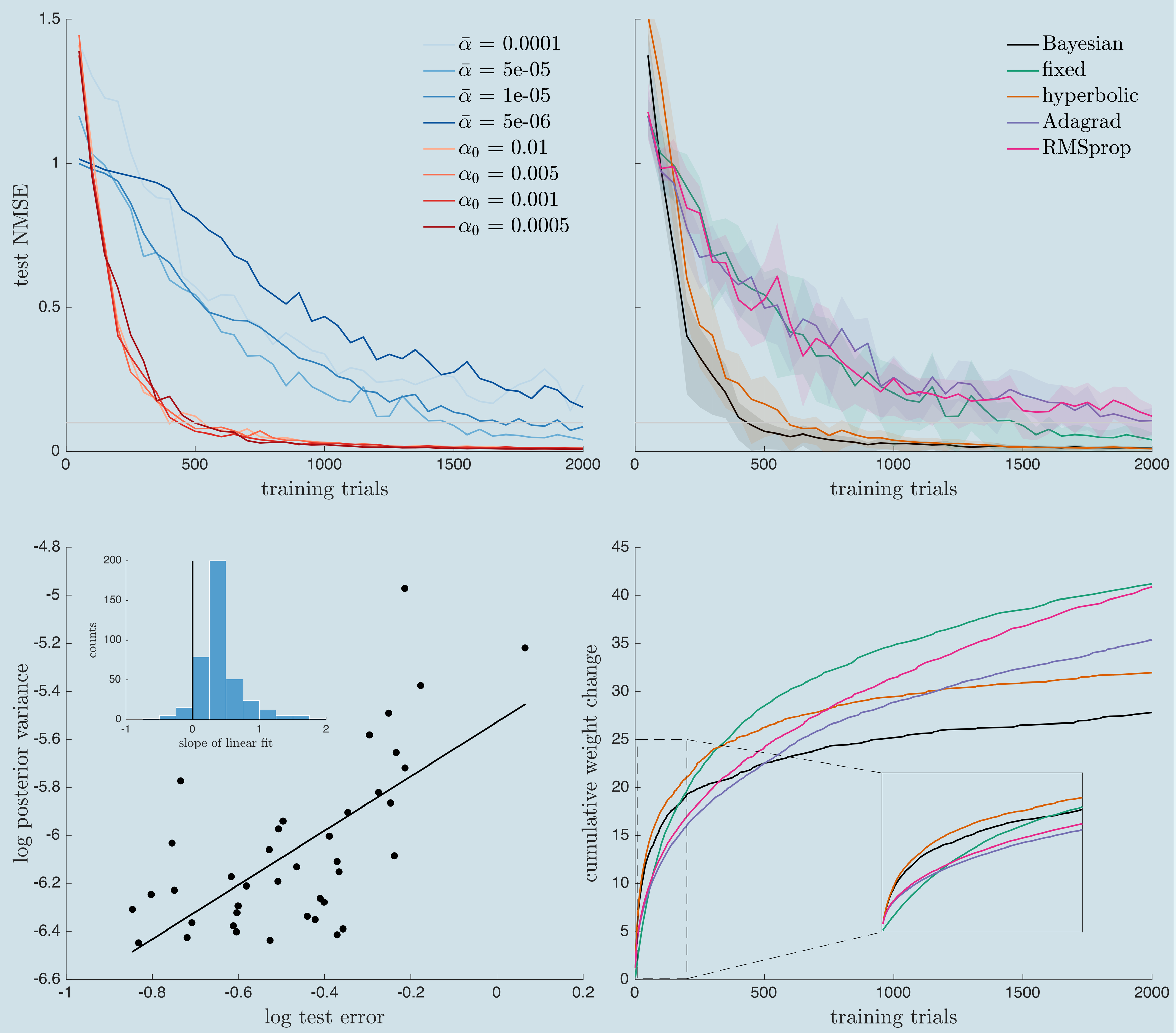
- Gaussian posterior yields simple learning rule:

$$w_i(t) = w_i(t-1) + \alpha_i(t)\delta(t)\phi(x_i(t))$$

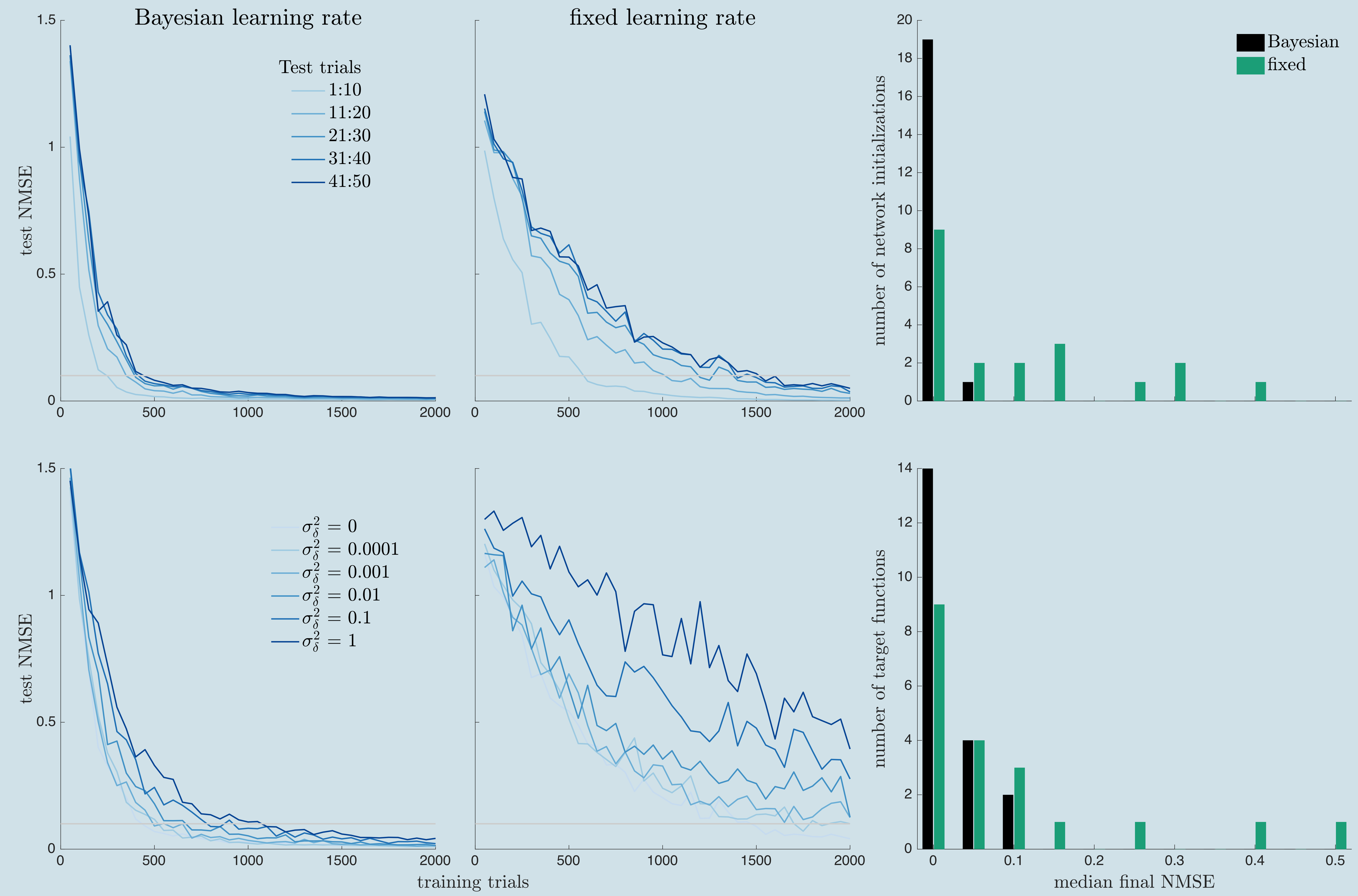
$$\alpha_i(t) = \frac{\sigma_i^2(t)}{\sigma_\ell^2} = \frac{1}{\frac{\sigma_\ell^2}{\sigma_i^2(0)} + \sum_{t'=1}^t \phi(x_i(t'))^2}$$

a delta rule with *synapse-specific* and *time-varying* learning rate

Faster learning



Improved 'generalization'



Conclusions

We find that a normatively derived learning rule implies a particular adaptive learning rate that improves stability of readout weights for reservoir computing.

The ***Bayesian plasticity hypothesis*** provides a general framework for formalizing normative principles of synaptic plasticity

- extension to Dale's law[1]
- learning recurrent weights[2]
- spike-based learning rules[3]
- heterosynaptic plasticity
- dendritic branching

References

1. Aitchison, L., Pouget, A., & Latham, P. E. (2014). Probabilistic Synapses. *arXiv preprint arXiv:1410.1029*.
2. Miconi, T. (2017). Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *Elife*, *6*.
3. Pfister, J. P., Toyoizumi, T., Barber, D., & Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural computation*, *18*(6), 1318-1348.

Funded by UCL Graduate Research Scholarship

jorge.menendez.15@ucl.ac.uk