Probabilistic Synapses

Aitchison et. al.

Two main concepts

Bayesian Plasticity - During learning, synapses take uncertainty into account

Synaptic Sampling - Weights are sampled from a probability distribution

Deterministic / Classical Model

PSP combine linearly

$$V(t) = \sum_i w_i(t) x_i(t) + \eta v(t)$$

Target weights and Synapse potential

$$V_{tar}(t) = \sum_i w_{tar}(t) x_i(t)$$

Feedback / error

$$\delta(t) = V_{tar}(t) - V(t) + \eta_\delta(t)$$

The delta rule to change the mean PSP

$$\Delta m_i = \alpha x_i \delta$$

Learning rate　Presynaptic term　Postsynaptic term

PSP: Post Synaptic Potential

Delta Rule

$$\Delta m_i = \alpha x_i \delta$$

Synapse 1 → certain about its weights and Synapse 2 → uncertain about its weights

Synapse 2 should have a larger impact of learning than synapse 1

Delta Rule

$$\Delta m_i = \alpha x_i \delta$$

Synapse 1 → certain about its weights and Synapse 2 → uncertain about its weights

Synapse 2 should have a larger impact of learning than synapse 1

Thus synapses keep track of probability distributions (log normal Probability distribution) -- Bayesian Plasticity

Delta Rule

$$\Delta m_i = \alpha x_i \delta$$

Synapse 1 → certain about its weights and Synapse 2 → uncertain about its weights

Synapse 2 should have a larger impact of learning than synapse 1

Thus synapses keep track of probability distributions (log normal Probability distribution) -- Bayesian Plasticity

Always gives positive values → so that the learning update does not change an excitatory synapse and turn it inhibitor

Optimal Learning Rule

$$\Delta m_i = \alpha x_i \delta$$

Synapse 1 → certain about its weights and Synapse 2 → uncertain about its weights

Synapse 2 should have a larger impact of learning than synapse 1

Thus synapses keep track of probability distributions (log normal Probability distribution) -- Bayesian Plasticity

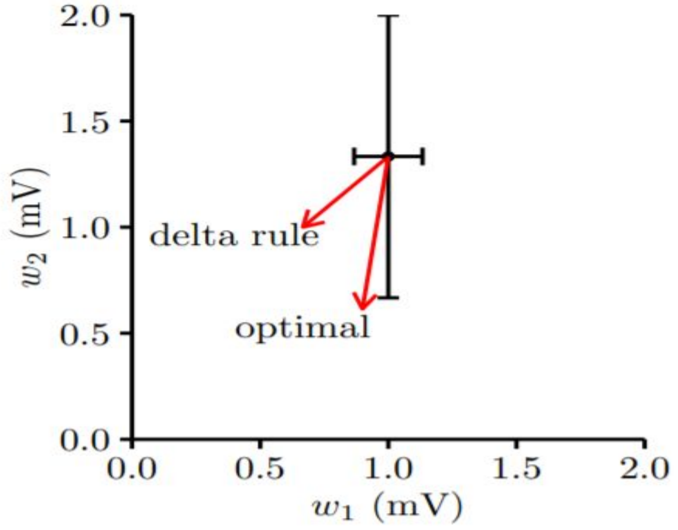
$$\Delta m_i = \alpha_i x_i \delta - \frac{1}{\tau}(m_i - m_{prior})$$

$$\Delta s_i^2 = -\alpha_i x_i^2 s_i^2 - \frac{1}{\tau}(s_i^2 - s_{prior}^2)$$

$$\alpha_i = \frac{s_i^2}{s_\delta^2}$$

Variable Learning rates - Bayesian Plasticity

Comparison of Delta and Optimal Learning Rule



Synaptic Sampling

PSP variability is a proxy for uncertainty - Synaptic Sampling Hypothesis

PSP variance = S_i^2

Synaptic Sampling

PSP variability is a proxy for uncertainty - Synaptic Sampling Hypothesis

PSP variance = S_i^2

The synapse uncertainty should fall as pre synaptic firing rate increases

$$\frac{\Delta m_i}{m_i} \propto \frac{1}{\sqrt{V_i}}$$

Combine bayesian Plasticity and Synaptic Sampling

The relative change in PSP amplitude amplitude is proportional to Synapse's uncertainty

$$\frac{\Delta m_i}{m_i} \propto \frac{PSPvariance}{PSPmean}$$

Normalized variability is inversely proportional to the firing rate

$$\frac{1}{\sqrt{V_i}} \propto \text{Normalized variability}$$

Feedback / $\boldsymbol{\delta}$

Supervised Learning $f(\delta) = \delta$

Supervised Learning with Binary feedback $f(\delta) = sign(\delta - \theta)$

Reinforcement Learning $f(\delta) = -|\delta|$

Unsupervised Learning - Since there is no feedback signal we need information from **x** and find $P(x|w_{tar}, V_{tar})$

Inference

Inference is a two step process:

1. Synapse incorporate new data following Bayes Theorem

$$P(\lambda_{tar,i} \mid D_i) = P(\lambda_{tar,i} \mid d_i, D_i(t-1)) \propto P(d_i|\lambda_{tar,i})\, P(\lambda_{tar,i}|\, D_i(t-1))$$

2. Synapse takes random changes of target weight into account

$$P(\lambda_{tar,i}(t+1) \mid D_i) = \int d\lambda_{tar,i}\; P(\lambda_{tar,i}(t+1)|\lambda_{tar,i})\; P(\lambda_{tar,i}|D_i)$$

However, the exact distribution is too complex to work with.

Inference Approximation

A Gaussian in the log-domain is chosen with mean **μ** and Variance ***σ²***

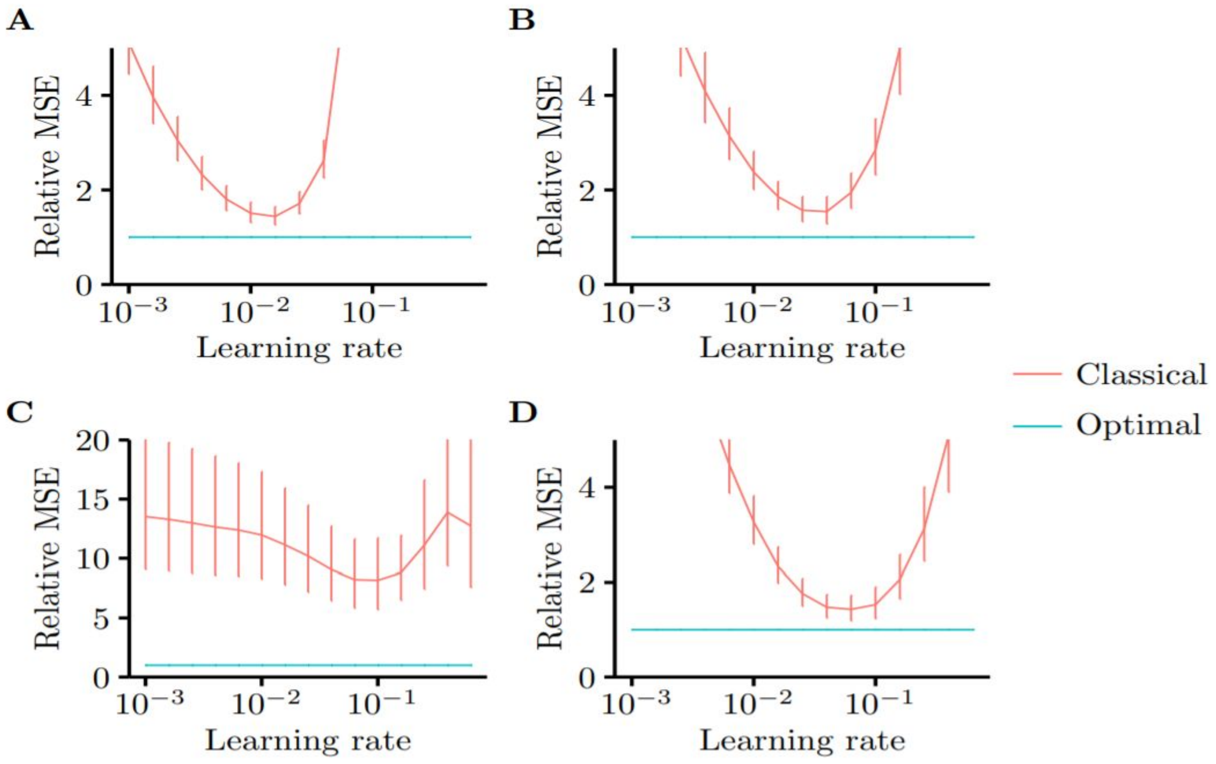
$$P(\lambda_{tar,i}(t+1) \mid D(t-1)) = \mathcal{N}(\lambda_{tar,i}\,;\mu_i\,,\sigma_i^2)$$

Corresponding mean ***m*** and variance ***s²*** over distribution ***w*** is:

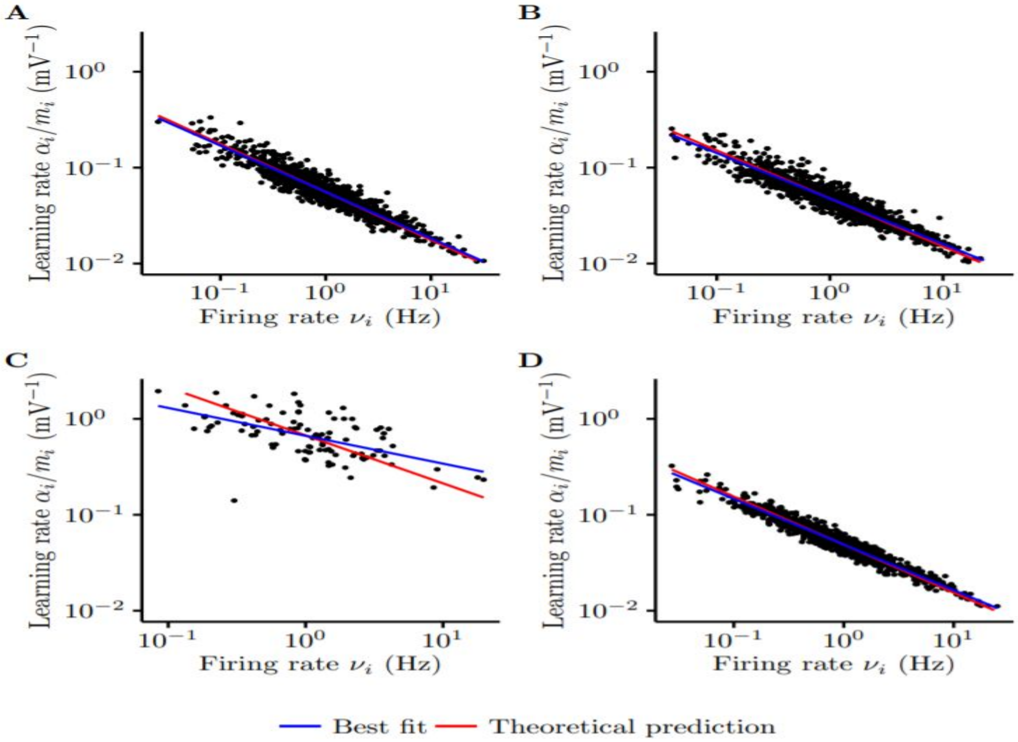
$$m_i \equiv E[w_{tar,i}|D(t-1)] = e^{\mu_i + \sigma_i{}^2/2}$$

$$s_i^2 \equiv Var[w_{tar,i}|D(t-1)] = (e^{\sigma_i^2} - 1)m_i^2 \approx \sigma_i^2 m_i^2$$

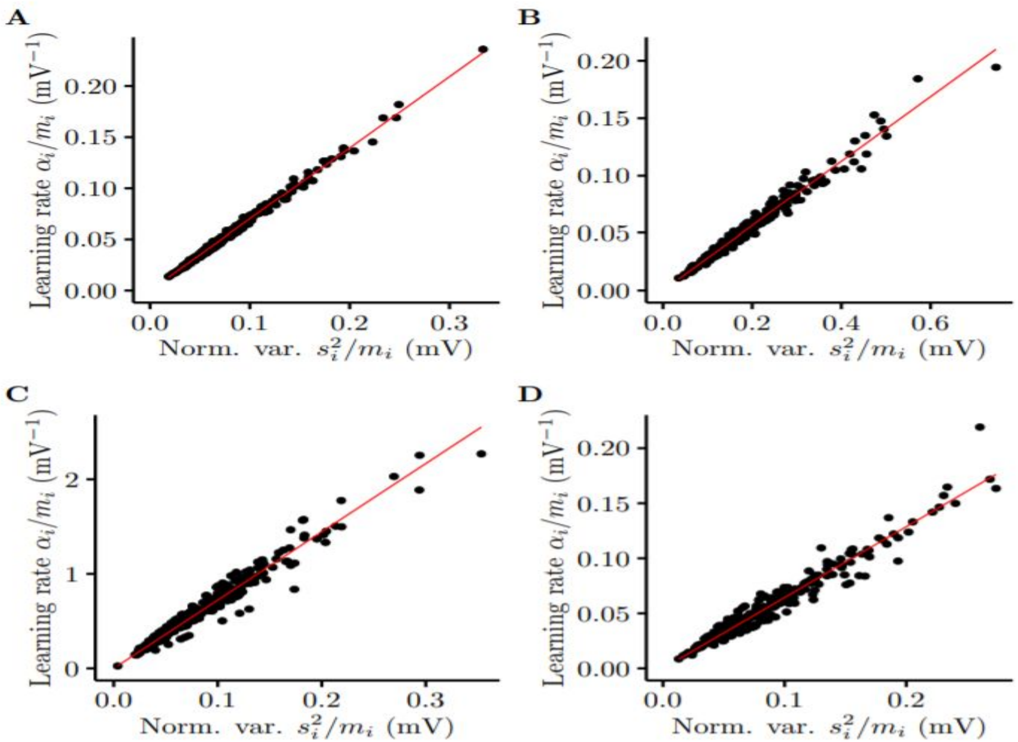
Classical vs Bayesian Learning



Normalized learning rate is inversely related to the square root of the firing rate.



Normalized learning rate is proportional to the normalized variability



Deep Learning perspective

- Neural Networks are deterministic in nature
- Uncertainty is necessary in safety critical application eg. Lung cancer prediction
- **Bayesian Neural Networks** (Wednesday class)