Upper Confidence Bounds in Basal Ganglia Model

Beren Millidge

September 24, 2022

Exploration-Exploitation Tradeoff with stochastic bandits

- In stochastic choice tasks, there is always uncertainty about the true distribution of rewards.
- This uncertainty prodocues an exploration-exploitation trade-off.
- It is not necessarily the case that the choice that seems the best actually is – you may have just gotten unlucky.
- You also need to explore other options, to reduce your uncertainty about their reward distribution.
- But this exploration has a cost – the opportunity cost of not selecting the best option

Upper Confidence Bound Algorithm

- Upper Confidence Bound (UCB) solves this problem extremely well for stochastic bandit tasks.
- Intuition is to give the estimated value of each choice an 'uncertainty bonus', effectively starting out optimistic
- As you sample a specific option more often, its uncertainty bonus decreases relative to the others, making it less attractive.

Upper Confidence Bounds in the Basal Ganglia

- Mathematically (under Gaussian assumptions), this is represented as:

$$Q^{UCB}(a) = Q(a) + \frac{c\sigma(a)}{\sqrt{(N(a))}} \tag{1}$$

- But we can write both $Q(a)$ and $\sigma(a)$ in terms of Go and No-Go neuron firing

$$Q = G - N \tag{2}$$

$$\sigma = \sqrt{\frac{\pi}{2}}(G + N) \tag{3}$$

Upper Confidence Bounds in the Basal Ganglia

- After some algebra, this lets us write,

$$Q^{UCB} = (1 + c\sqrt{\frac{\pi}{2N(a)}})G - (1 - c\sqrt{\frac{\pi}{2N(a)}})N \qquad (4)$$

- Effectively, we can implement the UCB uncertainty bonuses purely through modulating the weights of the Go and No-Go pathway.
- Increasing Go weights increases uncertainty bonus, and vice versa.
- Only additional information needing to be tracked is the number of actions sampled $N(a)$.
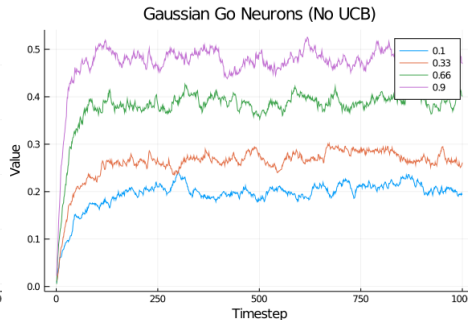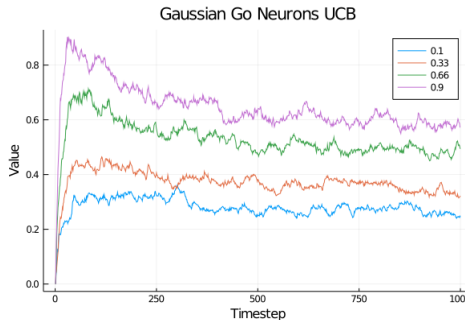
Simulation of Go Neuron Pathway
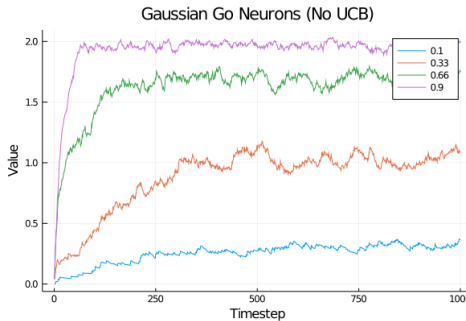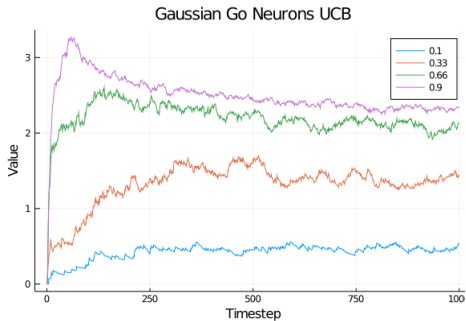
Simulation of No-Go Neuron Pathway
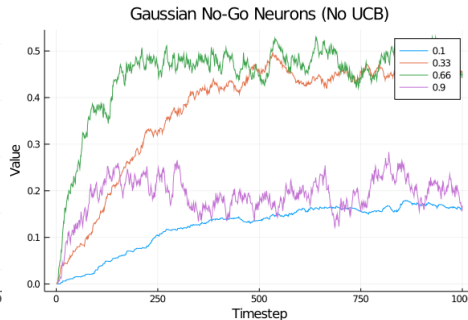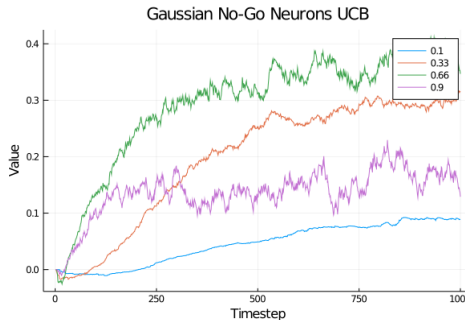
Simulation of Gaussian Go Neuron Pathway

Simulation of Gaussian No-Go Neuron Pathway