Understanding neural networks using stochastic simulation and information theory

Using stochastic simulation

During my MSc, I had an idea of using t-tests to select neurons responsive to stimuli, which would imply that neuronal responses followed a parametric distribution. The code I wrote, using t-tests, was able to select responsive neurons.

I also tried using t-tests on a small sample of prime numbers. That too returned a positive correlation. I was also able to obtain curves, showing a log-normal distribution for both these phenomena.

I am now looking to develop this further, and use Monte Carlo simulations to infer the statistical properties of neuronal responses and prime numbers. If successful, it would prove that both these are log-normal stochastic processes. The next step would be to look at both through the lens of random matrix theory.

The work on neurons relates to research carried out by Prof. Gyorgy Buzsaki, among others, who have given evidence for the log-normal distribution. Monte Carlo simulations could provide computational proof. This work could add to the growing support for the temporal coding theory. It could also contribute to our understanding of neural probe function and neural networks.

I also want to explore how the log-normal distribution of primes relates to the Erdos-Kac theorem, the prime number theorem and the work being done on gaps between primes.The distribution of prime numbers could possibly have implications for computational number theory and cryptography.

(code available here: https://github.com/sidkackar/MSc-code).

Reference: Buzsáki, G., Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci 15, 264–278 (2014).

<u>Using information theory</u>

Introduction

Information theory provides mathematically rigorous tools to quantify the precision of information transmission, setting theoretical limits on maximum information capacity (Borst and Theunissen 1999). In order to understand the transfer of information in the nervous system, we need to understand the relationship between a stimulus and the response it evokes in a single or network of neurons. In addition, we also need to understand how that stimulus is represented or coded in the neuron and how that representation relates to observed behavior. Neural coding is concerned with the study of these two aspects of information flow. I want to use our current knowledge of neural encoding and information theory to study how the brain perceives the statistical features of the input and encodes them to form representations of our world. Also, I want to study how the brain stores this information in the longer term i.e. how it learns from probabilistic stimuli. Ultimately, I want to contribute toward a unified understanding of perception and learning. This will require the combined efforts of electrophysiologists, molecular biologists, cognitive scientists and I want to contribute towards this problem from a computational/theoretical standpoint.

Information theory and neural coding

According to Borst and Theunissen (1999), the three central questions in the field of neural coding are to find out:

What is being encoded? For example, whether the information is coded in the amplitude of spikes or the change in amplitude?
How is it being encoded? The question of rate vs temporal coding has seen conflicting reports, with conventional studies supporting rate coding as standard (Adrian and Zotterman, 1926), but more recent studies reporting examples of temporal coding (Stein et al, 2005).
With what precision is it being encoded? What is the degree of variability in the responses?

Neuroscientists have traditionally addressed the first two questions by studying stimulus-response curves and changing the stimulus ensembles and response measures. Error bars were used to assess the variability. Information theory, on the other hand, provides mathematically rigorous tools to quantify the precision of information transmission, setting theoretical limits on maximum information capacity. The theory can be applied to study neural coding in the following ways:

Estimating maximum information capacity (channel capacity) of a neuron and the actual information transmitted, to quantify efficiency.
Compare the upper bound of information transmitted with that of an optimal linear model, to test for non-linearities and find out how a neuron transforms our data.
Determining the limiting temporal precision of code i.e. the minimal timescale in which information is contained, to find out if there is any information in the precise timings of the spikes.

To calculate maximal information transfer, a new quantity was introduced, called entropy. Entropy characterizes how many free states a variable can assume and what is the probability of each i.e. the variability. Entropy is the information needed to eliminate all uncertainty about a variable. In fact, information can be considered as a reduction in entropy. There are three common approaches used for estimating the mutual information between the stimulus and response, the direct method, which calculates the exact value of information and the upper and lower bound methods. Factors such as the experimental parameters and quality of stimulus ensembles determine our choice of method. Generally, a combination of all three is used as an optimal linear model.

Another area of application of information theory is population coding. Population coding is defined as a method to represent stimuli from multiple neurons. The response of each neuron has a probabilistic distribution over a set of stimuli, which is considered together with other neurons to characterize certain features of the input. Population analysis is reported to have several advantages over single neuron recordings in reducing uncertainty due to neuronal variability and the ability to represent different attributes of the stimulus simultaneously.

Quiroga and Panzeri (2009) have suggested using linear decoding algorithms for population recordings in conjunction with information-theoretic tools as mutual information gives a more comprehensive quantification of the information contained in a neuronal population, by evaluating the reduction of uncertainty about the stimuli that can be obtained from the neuronal responses. Representing uncertainty is important for decisions involving risks and may be fundamental for neuronal computations that take place in the presence of both sensory and neural noise.

The information bottleneck method

The aspect of information theory I am most interested in exploring is the information bottleneck method and its applications in neural coding and learning. Tishby et al (1999) derived the information bottleneck method in order to answer the question of how much relevant information survives a communication. They consider the rate-distortion theory, which they deem inadequate for this purpose, as it requires defining a rate-distortion measure. In the study, Tishby et al have identified the Kullback-Leibler (KL) divergence as the correct effective distortion measure using their principles. The KL divergence is a measure used to compare two probability distributions. It defines the penalty we pay in code length when trying to encode a source, defined in one probability distribution as if it were distributed according to another.

They define the relationship between the rate of information transfer and the preservation of relevant information. Suppose there is an input variable X with a hidden characteristic Y. In order to transfer relevant information about Y from X to the output state T, we try to maximize the mutual information between Y and T, which is bounded above by the mutual information between X and T (by the data processing theorem), which we are trying to minimize. What this means is that we are trying to preserve the maximum amount of relevant information about Y, while letting go of the unnecessary parts of X. This creates the “information bottleneck” in the T-X-Y Markov chain.

Applications of the information bottleneck

This principle has applications in understanding both neural coding and learning. In experiments on the H1 motion-sensitive neuron of the fly visual system, Schneidman et al (2002) tried to characterize the neural “dictionary” containing the relevant stimuli and their respective responses. They tried to quantify the compressibility of said dictionary i.e. they tried to find out which features of the stimuli are relevant. They found that the stimulus code was highly compressible meaning the neuron was sensitive to only a few significant features. This feature of compressibility was preserved in different flies.

The information bottleneck method also has applications in understanding deep learning as each layer of the deep neural networks can be treated as input and output points for the surrounding layers, requiring analysis of information compression by the bottleneck method. (Tishby and Zaslavsky 2015, Schwartz-Ziv and Tishby 2017) However, in a recent study, Saxe et al (2018) have argued against this method being used as an explanation of deep learning saying that the results of Schwartz-Ziv and Tishby are very specific to their setup and do not hold true in the general case. Therefore, this is an area of intense interest in both the neuroscience and machine learning communities.

Conclusion

Information theory informs us of the precision of the neural code, it can affect the choice of stimulus ensemble and it can also be used to determine the fit of our models of encoding for dynamical stimuli. In addition, tools derived from information theory have a tremendous scope of applications in varied areas within neuroscience.

References

- Adrian ED, Zotterman Y (1926). "The impulses produced by sensory nerve endings: Part II: The response of a single end-organ". J Physiol. 61 (2): 151–171.
- Borst A, Theunissen FE. “Information theory and neural coding”. Nat Neurosci. 1999 Nov;2(11):947-57.
- Saxe AM, Bansal Y, Dapello J, Advani M, Kolchinsky A, Tracey BD, Cox DD. “On the Information Bottleneck Theory of Deep Learning”. ICLR 2018.
- Schneidman E, Slonim N, de Ruyter van Steveninck R.R., Tishby N, and Bialek W. Analyzing neural codes using the information bottleneck method. NIPS, 2002.
- Stein RB, Gossen ER, Jones KE (May 2005). "Neuronal variability: noise or part of the signal?". Nat. Rev. Neurosci. 6 (5): 389–97.
- Tishby N, Pereira FC, Bialek W. “The Information Bottleneck Method”.The 37th annual Allerton Conference on Communication, Control, and Computing. 1999. pp. 368–377.
- Tishby N, Schwartz-Ziv R. “Opening the Black Box of Deep Neural Networks via Information”. ArXiv 2017.
- Tishby N, Zaslavsky N. “Deep learning and the information bottleneck principle”.2015 IEEE Information Theory Workshop (ITW) 2015.
- Quian Quiroga R, Panzeri S. “Extracting information from neuronal populations: information theory and decoding approaches.” Nat Rev Neurosci. 2009 Mar;10(3):173-85.