Quantifying the precision of decoding models for high-dimensional stimuli

Charles Zheng and Yuval Benjamini

December 19, 2016

Abstract

The analysis of encoding and decoding models is a common theme in both cell recording studies and in neuroimaging. A basic measure of the precision of a decoder is its accuracy at distinguishing k different stimuli. However, the fixed-k accuracy becomes insensitive beyond limited range of precision: low-precision decoders saturate at the chance accuracy $1/k$, while high-precision decoders saturate near perfect accuracy. On the other hand, the entire curve of accuracies for $k = 2, 3, ...$ provides a detailed and interpretable characterization of decoder performance. However, due to limited sampling, usually only a portion of the curve can be estimated: furthermore, it is unclear how to summarize the information in the curve by a single statistic. We show that under a high-dimensional limits, the mutual information becomes a sufficient statistic for reconstructing the entire accuracy curve, therefore suggesting the adoption of the mutual information as measure of decoder precision. Based on our theory, we develop a novel estimator of mutual information suited for high-dimensional settings (such as those found in neuroimaging), and also a procedure for extrapolating the accuracy curve to arbitrarily many stimuli.

1 Introduction

Both computational and cognitive neuroscience are concerned with understanding brain function: while computational neuroscience is concerned with understanding functionality at the level of the spiking behavior of individual

neurons and small neural populations, cognitive neuroscience tends to emphasize functionality at the level of macroscale regions of the interest in the brain. While the recording technologies, motivating questions, and analytical methodologies differ between the two subdisciplines, the conceptualization of brain functionality in terms of *encoding* and *decoding* models has been widely applied in both areas [24][17]. In computational neuroscience, cell recording experiments are conducted to determine whether spike trains have a temporal and/or correlational code [18][7], to examine how the neural code adapts to changes in stimulus distribution [3] and whether downstream neurons make use of higher-order correlations for decoding [19]. Meanwhile, in neuroimaging studies, functional MRI experiments are employed to model the receptive fields of early visual areas in the human brain [13], to examine the semantic encoding of words [16] or objects [10].

The dual perspectives of encoding and decoding originate naturally from the fact that in examining the link between brain activity and function, one can either start with brain activity on one end, or with external stimulation or behavioral observation on the other end. Starting by exposing the subject to sensory stimuli or prompting the subject to engage in particular motor tasks, one can search for areas in the brain which respond to the task: in other words, one can test to see which areas of the brain *encode* the given stimulus. In the other direction, one seeks to understand the functionality of a given brain region: in other words, how to *decode* brain activity in that region.

Formulation of encoding models is relatively straightforward, since one needs only to characterize the observed brain response to a given stimulus. One can further ask how to distinguish between signal and noise in the encoding mechanism [18], or in complex stimuli, seek a linearizing feature set which reveals the nature of the brain representation [17]. However, the establishment of complete decoding models is much less amenable to experiemental manipulation, since to exhaustively characterize the functionality of a neuron, one would have to know in advance the type of information it encodes. Early advances in decoding often depended on strokes of luck: Hubel [9] originally discovered the existence of neurons with orientation-sensitive receptive fields due to the vigorous response of a cell to the perfectly angled shadow of a glass slide that they were inserting into the ophthalmoscope. Yet, even now, the goal of completely characterizing the function of a given brain region remains a difficult task, with the most promising approach being a *reverse inference* procedure [23] which aggregates information from the

literature about activity-functionality relationships.

A more feasible goal is to establish the *precision* with which a neuron can decode a particular type of feature. This can be accomplished by first training an encoding model, and then inverting the encoding model using Bayes' rule to obtain a decoding model [20][24][17].

Measures of decoding precision can be used to support several different kinds of scientific inferences. When there exist multiple plausible encoding models–for instance, a model where stimulus information is encoded solely by average firing rate versus a model where inter-spike timings also carry information–the precision of the decoder can be used as a basis for deciding the best encoding model. For two encoding models with equal complexity, such as comparing two different types of receptive field models, the model with better decoding precision could be considered the more plausible model. In the case where a more complex encoding model is compared to a strictly simpler model–such as comparing a model with a temporal code versus a model only incorporating average firing rate, a substantial improvement in decoding precision for the more complex model is needed to demonstrate its validity, since in the null hypothesis where the simpler model is correct, the more complex model should still have approximately equal decoding performance.

Yet another application of decoding precision is to track the adaptivity of the neural code. Fairhall [3] recorded the output of a motion-sensitive neuron in a fly in response to a visual stimulus with changing angular velocity. Changing the variance of the stimulus results in rapid adaptation: the neural code starts adapting to the change in stimulus distribution within tens of milliseconds, which is reflected by an increased or decreased precision (as measured by mutual information) in resolving angular velocity to match the variance of the stimulus. More generally, comparisons of decoding precisions between different conditions can show how the encoded information increases or decreases across experimental conditions. Kayser [14] demonstrated how the mutual information between a sound stimulus and neurons in the auditory cortex increased when the subjects were also presented a matching visual stimulus (e.g. showing a picture of a lion roaring while playing the sound of a lion's roar.)

Differing types and parameterizations of stimuli naturally lead to differing measures of decoding precision. For stimuli which can be parameterized by a scalar x, the precision can be measured by the squared correlation coefficient R^2 [1]. However, the resulting measure of precision is not invariant

to scaling of the parameterization: for instance, the choice of whether to parameterize volume on an absolute scale or a logarithmic scale. The mutual information [25] between the stimulus and the predicted stimulus is invariant to the parameterization of the stimulus. Due to its invariance and a number of other properties, the mutual information is widely used to measure the precision of the neural code in cell recording studies, both for single-neuron decoding models [2] and for population coding models [24][12].

However, the difficulties of estimating mutual information in small samples has been widely recognized, with a large literature on bias correction methods [22][21]. Methods for bias correction have been developed for three different sample size regimes: the moderate-sample regime, where the number of observations is larger than the number of stimulus-response pairs[15][26][27], the undersampled regime, where the number of observations is less than the number of stimulus-response pairs [**?**], and a *stimulus-undersampled* regime, where only a small fraction of possible stimuli are sampled, but with a large number of observations for each of the sampled stimuli [5]. Nevertheless, even the bias-corrected estimates may be unusably inaccurate in problems of moderate dimensionality, since the cardinality of response space grows exponentially with the dimensionality. In such cases, alternative approaches for estimating the mutual information include the assumption of a parametric model [5], or usage of the maximum entropy principle to obtain bounds on the mutual information subject to the empirical moments of a certain order [11][6].

Perhaps due to the technical difficulties of estimating mutual information in high dimensions, mutual information has seldom been employed as a measure of decoding precision in neuroimaging studies, although it has been proposed for the purpose of bypassing the modelling of the hemodynamic response function for single-voxel analyses [4]. Instead, classification accuracy [8] is the dominant measure of precision for multivoxel decoding models.

References

[1] Larry F. Abbott. Decoding neuronal firing and modelling neural networks. *Quarterly reviews of biophysics*, 27(3):291–331, aug 1994.

[2] Alexander Borst and Frédéric E. Theunissen. Information theory and neural coding. *Nature Neuroscience*, 2(11):947–957, nov 1999.

[3] Adrienne L. Fairhall, Geoffrey D. Lewen, William Bialek, and Robert R. de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412(23):787–792, aug 2001.

[4] Galit Fuhrmann Alpert, Fellice T. Sun, Daniel Handwerker, Mark D'Esposito, and Robert T. Knight. Spatio-temporal information analysis of event-related BOLD responses. *NeuroImage*, 34(4):1545–1561, 2007.

[5] Michael C. Gastpar, Patrick R. Gill, and Frédéric E. Theunissen. Anthropic correction of information estimates. *Proceedings - 2009 IEEE Information Theory Workshop on Networking and Information Theory, ITW 2009*, 56(2):152–155, 2009.

[6] Amir Globerson, Eran Stark, Eilon Vaadia, and Naftali Tishby. The minimum information principle and its application to neural code analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3490–5, mar 2009.

[7] N G Hatsopoulos, C L Ojakangas, L Paninski, and J P Donoghue. Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 95(26):15706–11, dec 1998.

[8] James V. Haxby, Andrew C. Connolly, and J. Swaroop Guntupalli. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, 37(1):435–456, jul 2014.

[9] David H. Hubel. Evolution of ideas on the primary visual cortex, 19551978: A biased historical account. *Bioscience Reports*, 2(7):435–469, 1982.

[10] Alexander G. Huth, Shinji Nishimoto, An T. Vu, and Jack L. Gallant. A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, 76(6):1210–1224, 2012.

[11] Robin A A Ince, Rasmus S Petersen, Daniel C Swan, and Stefano Panzeri. Python for information theoretic analysis of neural data. *Frontiers in neuroinformatics*, 3:4, 2009.

[12] Robin A.A. Ince, Riccardo Senatore, Ehsan Arabzadeh, Fernando Montani, Mathew E. Diamond, and Stefano Panzeri. Information-theoretic methods for studying population codes. *Neural Networks*, 23(6):713–727, 2010.

[13] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(March):352–355, 2008.

[14] Christoph Kayser, Nikos K. Logothetis, and Stefano Panzeri. Visual Enhancement of the Information Representation in Auditory Cortex. *Current Biology*, 20(1):19–24, 2010.

[15] Miller. Note on the bias of information estimates. *Information Theory in Psychology: Problems and Methods*, 1955.

[16] Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, 320(5880), 2008.

[17] Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fMRI. *NeuroImage*, 56(2):400–410, 2011.

[18] Israel Nelken, Gal Chechik, Thomas D Mrsic-Flogel, Andrew J King, and Jan W H Schnupp. Encoding Stimulus Information by Spike Numbers and Mean Response Time in Primary Auditory Cortex. *Journal of Computational Neuroscience*, 19:199–221, 2005.

[19] Masafumi Oizumi, Toshiyuki Ishii, Kazuya Ishibashi, Toshihiko Hosoya, and Masato Okada. Mismatched Decoding in the Brain. *Journal of Neuroscience*, 30(13):4815–1826, 2010.

[20] Mike W. Oram, Peter Földiák, David I. Perrett, Mike W. Oram, and Frank Sengpiel. The 'Ideal Homunculus': decoding neural population signals. *Trends in Neurosciences*, 21(6):259–265, 1998.

[21] Liam Paninski. Estimation of Entropy and Mutual Information. *Neural Computation*, 15(6):1191–1253, 2003.

[22] Stefano Panzeri, Riccardo Senatore, Marcelo A. Montemurro, and Rasmus S. Petersen. Correcting for the Sampling Bias Problem in Spike Train Information Measures. *Journal of Neurophysiology*, 98(3), 2007.

[23] Russell A. Poldrack. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2):59–63, 2006.

[24] Rodrigo Quian Quiroga and Stefano Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews. Neuroscience*, 10(3):173–185, 2009.

[25] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, jul 1948.

[26] S. P. Strong, Roland Koberle, Rob R. de Ruyter van Steveninck, and William Bialek. Entropy and Information in Neural Spike Trains. *Physical Review Letters*, 80(1):197–200, jan 1998.

[27] Alessandro Treves and Stefano Panzeri. The Upward Bias in Measures of Information Derived from Limited Data Samples. *Neural Computation*, 7(2):399–407, 1995.