

Quantifying the precision of decoders 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 $k = 2, 3, \dots$ 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 [11][7]. In computational neuroscience, cell recording experiments are conducted to determine whether spike trains have a temporal and/or correlational code [8][2], to examine how the neural code adapts to changes in stimulus distribution [1] and whether downstream neurons make use of higher-order correlations for decoding [9]. Meanwhile, in neuroimaging studies, functional MRI experiments are employed to model the receptive fields of early visual areas in the human brain [5], to examine the semantic encoding of words [6] or objects [4].

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 [8], or in complex stimuli, seek a linearizing feature set which reveals the nature of the brain representation [7]. However, the establishment of complete decoding models is much less amenable to experimental 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 [3] 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 [10] 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 to obtain a decoding model [7].

References

- [1] 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.
- [2] 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.
- [3] David H. Hubel. Evolution of ideas on the primary visual cortex, 1955-1978: A biased historical account. *Bioscience Reports*, 2(7):435–469, 1982.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fMRI. *NeuroImage*, 56(2):400–410, 2011.

- [8] 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.
- [9] Masafumi Oizumi, Toshiyuki Ishii, Kazuya Ishibashi, Toshihiko Hosoya, and Masato Okada. Mismatched Decoding in the Brain. *Journal of Neuroscience*, 30(13):4815–1826, 2010.
- [10] Russell A. Poldrack. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2):59–63, 2006.
- [11] Rodrigo Quian Quiroga and Stefano Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews. Neuroscience*, 10(3):173–185, 2009.