Tomography for Neural Networks

October 14, 2023

Abstract

The characterization, verification and validation of quantum devices (QCVV) has developed an extensive theory of *tomographic* methods, wherein quantum circuits are constructed from a collection of primitive gates, and these circuits are run many times to collect statistics. Using these data, it is possible to come up with estimates for an underlying model of the system, which at its most basic is the gate set. (This can be augmented with preparation errors and measurement errors, as well as potentially non-Markovian error models.) Recent work by Zou et al. has introduced a linear artificial tomography to achieve similar goals. We discuss the connection in this note.

We will focus on the neural activity for a concept. This corresponds to a vector of non-negative numbers, where each vector is labeled by a stimulus. The neural sequence of length N is a collection of N experiments (each $M \leq N$ length subsequence) which is a HMM given by the neural network architecture.

Consider the amount of ¡concept¿ in the following: ¡stimulus¿ The amount of ¡concept¿ is

In particular, the set A_c of neural activity for a concept is a collection of vectors, for a collection of stimuli S. In GST, for any collection of experiments there is a vector of outputs, and hence we can imagine the particular choice of SPAM for each experimental sequence (often used for process tomography) as the set of stimuli.

USER: ¡instruction¿ ¡experimental/reference prompt¿ ASSISTANT: ¡output¿

What if the read on the assistant data is corrupted? "Pretend you are dishonest" and it acts differently.