Neural representations that are good for both generalization and discrimination

Omri Barak, Mattia Rigotti and Stefano Fusi

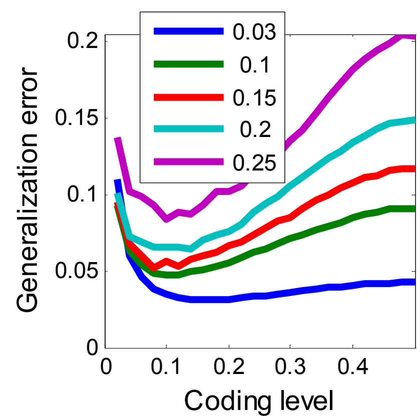
How do we tell whether a neural representation is good or bad? The answer depends on the input statistics, the task at hand and the readout. Previous work, mainly in early sensory areas, focused on the amount of information about a stimulus contained in the neural representation. Here we took a different perspective and evaluated a neural representation by considering the dynamics of a generic cortical circuit. This approach led us to test the classification performance of a linear readout from a population of input neurons that encode a few different noisy sources of information. Our analysis revealed that input neurons have to respond to mixtures of the information sources in order to enable classification. One efficient way to achieve these response properties is to mix different sources of information with randomly connected neurons (RCNs). Under the assumption that the output neuron reads out the RCNs, we derived a formula for the classification performance of noisy inputs (generalization error). The performance depends on the "discrimination factor", expressing how much the population activity changes when only one of the information sources is altered; and on the "generalization factor", expressing the change when none of the sources are altered, but different noisy versions are presented. Specifically we explored the effect of the population coding level on the tradeoff between these factors, and show that a coding level of about 0.1 is optimal for many different cases. The advantage of optimal coding is greater for higher levels of noise in the inputs. Our results provide a possible explanation for the abundance of mixed selectivity found in neural recordings, and for the coding level observed in many areas. Furthermore, we provide a prescription to measure components of the generalization-discrimination tradeoff from neural data.
Funding: DARPA SyNAPSE. Gatsby, Kavli and Sloan-Swartz Foundations.

Additional information

The theoretical framework we use is an extension of [1] to multiple sources of information, each represented by a population of binary neurons. Each information source can take one of several states. A second layer of Randomly Connected Neurons (RCNs) receives fixed random connections from all sources, and each neuron is active if its input passes a given threshold θ (this determines the coding level which is the fraction of stimuli eliciting a response in a given RCN). A Perceptron is trained to classify randomly different combinations of noisy realizations of the information sources.

The two main results are:

1) When trying to integrate two sources of information, one having m possible states and the other having n possible states, we face a classification problem of mn patterns in $m+n$ dimensions. In general this will not be linearly separable, and hence the need for mixing the sources of information in a non-linear way. The problem is even more severe for more than two sources of information.
2) When presenting $p=mn$ noisy patterns, with a probability q of flipping any bit in the input, the test error can be approximated by $\frac{1}{2} erfc\left(\sqrt{\frac{\gamma(\theta,q)}{\sigma^2(\theta,q,p)}\frac{N_{RCN}}{p}}\right)$. γ measures the non linearity of the input to RCN transformation, can be measured from data, and generally increases with coding level. σ^2 measures the variability in RCN space that affects classification, is roughly equal to the inter trial variability, is very weakly dependent on p, and increases with coding level.



Dependence of test error for different noise levels q. Note that the steepness increases with the noise, and the optimal coding level slightly decreases.

[1] M. Rigotti, D. B. . Rubin, X. J. Wang, and S. Fusi, "Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses," *Frontiers in Computational Neuroscience*, vol. 4, 2010.