Population Coding Of Categories

Laurent Bonnasse-Gahot[†*] **and Jean-Pierre Nadal**[†‡]

[†]Centre d'Analyse et de Mathématique Sociales, UMR 8557, CNRS-EHESS
[‡]Laboratoire de Physique Statistique, UMR 8550 CNRS-ENS-Paris 6-Paris 7
[*]lbg@ehess.fr

Abstract

This work deals with the analytical study of coding a discrete set of categories by a large assembly of neurons. We consider population coding schemes, which can also be seen as instances of *exemplar models* proposed in the literature to account for phenomena in the psychophysics of categorization. We quantify the coding efficiency by the mutual information between the discrete categories and the neural code and characterize the properties of the most efficient codes in the limit of a large number of coding cells. One key result is that the highest stimulus-discriminating parts of the neuronal tuning curves should be placed in the transition regions between categories in stimulus space.

1. Methods

1.1 General Framework

Processing chain:

$$\mu \rightarrow \mathbf{x} \rightarrow \mathbf{r} \rightarrow \widehat{\mu}$$

- $\mu = 1, \ldots, M$ categories
- Probabilities of occurrence $q_\mu \geq 0$ $(\sum_\mu q_\mu = 1)$
- Input (sensory) space $\mathbf{x} \in \mathbb{R}^K$
- A category $\equiv P(\mathbf{x}|\mu)$
- Neural response $\mathbf{r} = \{r_1, \ldots, r_N\}$
- $\widehat{\mu}$: category estimate (decoding)

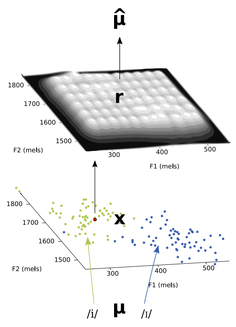


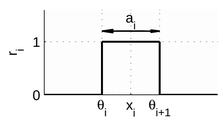
Figure 1: *Example of the feedforward processing*

1.2 Population Coding

- population of N cells
- activity r_i of cell i given by $P_i(r_i|\mathbf{x})$
- no correlations (given $\mathbf{x}$): $P(\mathbf{r}|\mathbf{x}) = \prod_{i=1}^N P_i(r_i|\mathbf{x})$
- each cell i is stimulus selective
- tuning curve (mean response): $f_i(\mathbf{x}) = R_i\, f(\mathbf{x}, \mathbf{x_i}, \mathbf{a_i})$ where
  - $\mathbf{x}_i$: preferred stimuli
  - $\mathbf{a}_i$: width of the receptive field
  - R_i : maximal rate

one dimensional case: $f_i(x) = R_i\, f(\frac{x - x_i}{a_i})$

2.1.2 Sharp Tuning Curves



$$r_i = \begin{cases} 1 \text{ if } x \in [\theta_i, \theta_{i+1}] \\ 0 \text{ otherwise} \end{cases}$$

Hypothesis: a_i small.

Mutual Information: Sharp Tuning Curve (1-d)

$$I(\mu, x) - I(\mu, \mathbf{r}) = \frac{1}{2} \sum_i a_i\, p(x_i)\, F_{\text{cat}}(x_i) \times \frac{a_i^2}{12}$$

where

- $\frac{a_i^2}{12}$ is the minimal variance of any $\widehat{x}(\mathbf{r})$
- $F_{\text{cat}}(x_i) = \sum_{\mu=1}^M P'(\mu|x_i)^2 / P(\mu|x_i)$

2.2 Short Time Limit

The previous results concern regimes of reasonably well-defined classes and sufficiently large signal-to-noise ratio. We have also studied noisy/short-time processing.

At most one cell emits one spike during a short time window $[0, \tau]$:
$P_i(r_i = 1|\mathbf{x}) = f_i(\mathbf{x})\ \tau$
$P_i(r_i = 0|\mathbf{x}) = 1 - f_i(\mathbf{x})\ \tau$
with $0 < f_{min} \leq f_i(\mathbf{x}) \leq f_{max}$ for every input $\mathbf{x}$ and every cell i. Assuming $\tau N f_{max} \ll 1$, one gets

$$I(\mu, \mathbf{r}) = \tau \sum_i \bar{f}_i \sum_{\mu=1}^M q_{i,\mu}\ \ln \frac{q_{i,\mu}}{q_\mu}$$

where $\bar{f}_i = \int d^K\mathbf{x}\ P(\mathbf{x})\ f_i(\mathbf{x})$ is the mean rate of cell i averaged over all possible inputs, and $\bar{f}_i^\mu = \int d^K\mathbf{x}\ P(\mathbf{x}|\mu)\ f_i(\mathbf{x})$ its mean rate conditional to the category μ.

The mutual information is maximized by maximizing the Kullback divergence between the probabilities $\{q_{i,\mu}, \mu = 1, ..., M\}$ and $\{q_\mu, \mu = 1, ..., M\}$: each cell must be as specific as possible to one and only one category. This implies that, in an optimized code, most cells will have a receptive field avoiding any domain in the input space where categories overlap.

1.3 Goals

- quantify the **coding efficiency** of the neural population with respect to the classification task at hand
- characterize the **optimal neural code**

Coding efficiency is quantified by the **mutual information** $I(\mu, \mathbf{r})$ between the categories μ and the neural response $\mathbf{r}$, defined by:

$$I(\mu, \mathbf{r}) = \sum_{\mu=1}^M q_\mu \int d^N\mathbf{r}\ P(\mathbf{r}|\mu)\ \ln \frac{P(\mathbf{r}|\mu)}{P(\mathbf{r})}$$

where $P(\mathbf{r})$ is the probability density function of $\mathbf{r}$.

2. Results

2.1 Information Content for Large but Finite N

2.1.1 Smooth Tuning Curves

Hypotheses:

- smooth tuning curves
- for any neural activity the maximum likelihood $\widehat{x}(\mathbf{r})$ is well defined and unique
- Large N limit $(N \gg K)$

Mutual Information: Smooth Tuning Curve (K-d)

$$I(\mu, \mathbf{x}) - I(\mu, \mathbf{r}) = \frac{1}{2} \int d^K\mathbf{x}\, p(\mathbf{x}) F_{\text{cat}}(\mathbf{x}) : F_{\text{code}}(\mathbf{x})^{-1}$$

where $\mathbf{x} = \{x_k, k = 1, ..., K\}$, and:

- $\left[F_{\text{code}}(\mathbf{x})\right]_{kl} = -\int d^N\mathbf{r}\, P(\mathbf{r}|\mathbf{x}) \frac{\partial^2 \ln P(\mathbf{r}|\mathbf{x})}{\partial x_k \partial x_l}$
  is the Fisher information matrix characterizing the sensibility of $\mathbf{r}$ with respect to small variations of $\mathbf{x}$
- $\left[F_{\text{cat}}(\mathbf{x})\right]_{kl} = -\sum_{\mu=1}^M P(\mu|\mathbf{x}) \frac{\partial^2 \ln P(\mu|\mathbf{x})}{\partial x_k \partial x_l}$
  is the Fisher information matrix characterizing the sensibility of μ with respect to small variations of $\mathbf{x}$
- $\forall A, B \in \mathcal{M}_K(\mathbb{R}),\ A : B = tr(A^\top B) = \sum_{k,l} A_{kl} B_{kl}$

Expectations on the optimal neural code.
We assume $P(\mu|x)$ to have a S-shape. This entails that $|\partial P(\mu|x)/\partial x|$ is the greatest at the boundary between categories. $F_{\text{cat}}(x)$ is therefore typically the greatest in these regions. $F_{\text{code}}(x)$ for a given cell is the highest at the flanks of the cell where the slope of the tuning curve is the steepest. [Seung and Sompolinsky, 1993]

As $N < \infty$, an optimized code leads to:

- a larger value of $F_{\text{code}}(x)$ at the class boundaries
- more cells coding for boundaries, *ie*
  - their steepest slope will be located in these regions
  - between categories cells will have a sharper tuning curve
  - typically, more cells are therefore expected at the boundary

Numerical Illustration (1-d)

- two Gaussian categories, centered at 0 and 1
- bell-shaped tuning curves: $f_i(x) = \exp\left(-\frac{(x - x_i)^2}{2a_i^2}\right)$
- (A) Optimal tuning curves obtained numerically.
- (B) As a function of x, difference between the Fisher information $F_{\text{code}}(x)$ for the optimal code, and the one for the equidistributed distribution of preferred stimuli

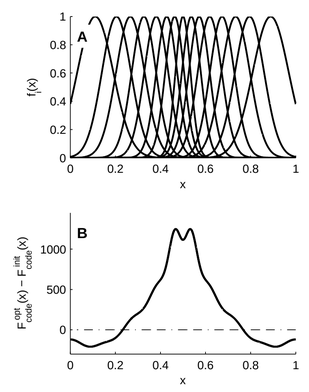


Figure 2: *1d numerical optimization*

Numerical Illustration (2-d)

- two Gaussian categories, centered at $(-2, 0)$ and $(2, 0)$
- bell-shaped tuning curves
- gray dashed contour: $P(\mathbf{x})$
- dashed circles: initial cells
- solid ellipses: configuration obtained from numerical optimization

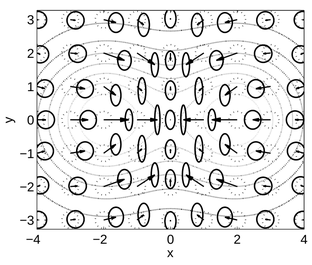


Figure 3: *2d numerical optimization* $(\mathbf{x} = (x, y))$

3. Discussion

3.1 Perceptual Consequences

- discriminability of stimuli x and $x + \delta x$ [Seung and Sompolinsky, 1993]:
  $$d' = |\delta x| \sqrt{F_{\text{code}}(x)}$$
- the higher the Fisher information $F_{\text{code}}(x)$ the higher the discriminability d'

If the code is optimized:

- F_{code} will typically be the greatest at the boundary
- higher cross-category than within-category discriminability
- Categorical Perception (CP) [Harnad, 1987]

Our results show that optimal coding underlies CP: **CP is a necessary byproduct of the minimization of misclassification probability.**

3.2 Category Learning and the Inferotemporal Cortex

The Inferotemporal (IT) cortex of the monkey is known as a site

- for object recognition and classification
- where population coding is a strategy widely used [e.g. Young and Yamane, 1992, Vogels, 1999]
- where perceptual learning implies neural modifications [Kobatake et al., 1998]

$\Rightarrow$ best candidate for testing our predictions

Experimental study of Freedman et al. [2003]:

- continuous set of morphed visual stimuli interpolating between cats and dogs
- two monkeys trained on a classification task (cats/dogs)
- almost half of all recorded IT neurons have preferred stimuli located at the class boundary

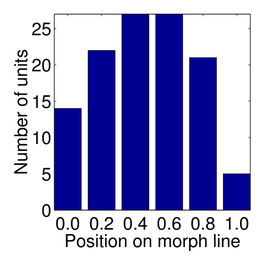


Figure 4: *Distribution of preferred stimuli for recorded IT neurons [Knoblich et al., 2002]*

3.3 Future Work

- study of the case of non-abutting categories
- generalization of the result to high-dimensional spaces
- consideration of noise correlations
- characterization of learning mechanisms that aim at attaining an optimal code, for both supervised and unsupervised learning of categories
- experimental investigations

Acknowledgments

This work is part of a project "Acqlang" supported by the French National Research Agency (ANR-05-BLAN-0065-01). LBG acknowledges a fellowship from the DGA. JPN is a CNRS member.

References

D. Freedman, M. Riesenhuber, T. Poggio, and E. Miller. *J. Neurosci.*, 15:5235–5246, 2003.

S. Harnad, editor. New York: Cambridge University Press, 1987.

U. Knoblich, D. Freedman, and M. Riesenhuber. *AI Memo 2002-007. Cambridge, MA: MIT AI Laboratory*, 2002.

E. Kobatake, G. Wang, and K. Tanaka. *J. Neurophysiol*, 80:324–330, 1998.

H. S. Seung and H. Sompolinsky. *PNAS*, 90:10749–10753, 1993.

R. Vogels. *Eur. J. of Neurosci*, 11:1239–1255, 1999.

M. Young and S. Yamane. *Science*, 256:1327–1330, 1992.

http://www.lps.ens.fr/∼Bonnasse-Gahot/

- L. Bonnasse-Gahot and J.-P. Nadal. Neural Coding of Categories. Submitted to *Journal of Computational Neuroscience*, May 2007. (downloadable preprint)
- L. Bonnasse-Gahot and J.-P. Nadal. From Exemplar Theory to Population Coding and Back – An Ideal Observer Approach. *Proc. of the workshop Exemplar-Based Models of Language Acquisition and Use, ESSLLI 2007, Dublin, Ireland, 6-17 Aug., 2007.*