Neuroscience, cognitive science and machine learning

Konrad Kording (with inspiration by Peter Dayan)
Level 3 Marr: Implementation

- In our brain - there are just spikes
- Neither suggestive of generative nor discriminative algorithm of learning
Neurons generate spikes

- Normally we do not care about generative model - just performance for decisions
- But we do want to understand the brain
- Which means understand the model which generates spikes
- Understand structure, function and objectives
It is a big problem

- Currently we can only record from a small percentage of neurons at a time
- and yet we want to understand how it works

$10^{11}$ Neurons
Hope #1

- Brain consists of areas
- Within each area neurons are essentially all identical
- Apart from small number of parameters
- Strong analogy to computer programs, e.g. neural networks
- Map how activity relates to outside world
Brain consists of groups of similar neurons

Neurons within each group are essentially identical

While their relation to the outside world may be complicated their interactions may be simple

Therefore, recording from each of the groups is enough
Hope #3

- Neurons may all be different from one another
- But their learning rules may be identical
- If we know the statistics of inputs we can predict distributions
- If we understand how they learn we might understand how it works
Hope #4: New technologies

- New technologies accelerate rapidly
- Record from all Neurons at the same time
- And then?
A well-known Law

* My phone = fastest computer on the planet in 1980
Implications

- Focus not on speed but on scaling
- O(n) notation
- What is O(n), O(n log n) etc?
A Much Less known Law

Simultaneously Recorded Neurons

Publication Date

N=56

Doubling Time: 7.4 ± 0.4 years
Implications

- Analyze scaling behavior
- Computer time per neuron (no worries)
- Information learned per neuron!
Outline of talk

- Understand how neurons relate to the outside world
- How neurons represent uncertainty
- How neurons relate to one another
Part 1: How neurons relate to the outside world
Experimental studies

- Vary the stimulus
- Vary the behavioral demands
- Measure behavior
- Measure neural signals
Motor Tuning in M1

Slide adapted from Dayan

Orientation Tuning in V1

Slide adapted from Dayan

Hubel & Wiesel (1968)
Disparity Tuning in V1

Slide adapted from Dayan

Poggio & Talbot (1981)
Spatial Tuning

Slide adapted from Dayan

entorhinal cortex

hippocampus

Moser et al
Bygone Actress Tuning

Slide adapted from Dayan
Bayesian tuning curve analysis

- We want to understand p (tuning curve | spikes)
- Markov-Chain-Monte-Carlo
- Cronin et al 2010
Part 2: How neurons may represent uncertainty
Many theories

- Brain has no reason to use code that Konrad can well understand

A. Low uncertainty
   - High uncertainty
   - Gain encoding (e.g. Ma et al)

B. Tuning width
   - Covariate, e.g. direction
   - Separate population (e.g. Schultz et al)

C. Probability change
   - Covariate, e.g. direction
   - (e.g. Deneve et al)

D. Spatiotemporal variability
   - Histogram (e.g. Hinton, Hoyer)
Distributed representation

- Distribution across neurons represents uncertainty
- Different variables represented by different populations
- + fast
- - requires many neurons

Fiser et al 2010
Review characteristics of neural tuning curves and noisiness probability distributions by this scheme are determined by the becomes approximate. Characteristics of the family of representable exponentially large and with fewer neurons the representation number of neurons needed in an exact PPC representation would be number of its variables, a drawback of such schemes is that the specify a multivariate distribution scales exponentially with the representational schemes particularly suitable for real-time inference determining its parameters, making PPCs and other parametric neurons in PPCs provide a complete description of the distribution by encoded the elements of the mean vector and covariance matrix of a highly unfeasible version of this scheme would be if different neurons the true distribution that needs to be represented (purple colormap). In this example, (populations 1 and 2, colored circles) determine a probability distribution over the represented variables (top right panel, contour lines), which Table I. Comparing characteristics of the two main modeling approaches to probabilistic neural representations Figure Ia. A simple but multimodal distributions critical factor in accuracy of encoding a distribution Representable distributions – deterministic Stochastic (self-consistent) Number of neurons needed for representing Number of dimensions Scales exponentially with the represented at any time Complete, the whole distribution is – parametric form Must correspond to a particular encoded by samples Distribution across neurons represents uncertainty Different variables represented by different populations + few neurons necessary - slow

Fiser et al. TICS 2010
Part 3: Reverse engineering the way neurons interact
Why model interactions?

N=56
Doubling Time:
7.4 ± 0.4 years
A generative model of spikes

\[ \lambda_i(t) = \exp \left( \sum_{i=1}^{N} \sum_{\Delta t=1}^{T} W_{i,\Delta t} s_i(t - \Delta t) \right) \]

- Biological interpretation
- Expressive power
- Real reason why we use this
- MAP estimation of weights
Explaining away

Cross Correlations

Functional Connections

<table>
<thead>
<tr>
<th>Triggers</th>
<th>Targets</th>
<th>Triggers</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>150</td>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td>250</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>150</td>
<td>350</td>
<td>400</td>
<td>150</td>
</tr>
<tr>
<td>200</td>
<td>450</td>
<td>500</td>
<td>200</td>
</tr>
</tbody>
</table>

A B C

Time (s)

A B C

Time (s)
Results from real neurons

**Figure 6** Reconstructing spatio-temporal kernels from simulated data. A typical reconstruction of the interactions in a simulated 4-cell network at 1 ms resolution. Cross-validated log-likelihood fits (N=1000) and correlation coefficients (N=.01) between reconstructions and ground truth connectivity for 40 cell networks. Error bars denote SEM.

**Figure 6 -6** Reconstructing spatio-temporal kernels from real data. Typical reconstructions for a subset of cells calculated from two segments fired and black. MAP estimates are shown in dark, ML estimates in light. The cross-validated log-likelihood, as well as fraction of non-zero connections, as a function of the hyperparameter b. The average correlation coefficient between MAP estimates from one segment and ML estimates from other segments. The connections themselves have similar properties to those found in rat hippocampal cells [14, 15]. Neurons interact with themselves with a refractory period followed by an excitatory rebound, and other connections, though fairly rare, tend to be weakly excitatory. Since ML estimates have no constraints on the smoothness of kernels, neurons with low spike rates tend to have especially noisy interactions. Using priors, these spurious connections are set to zero. The spatio-temporal kernels and connection weights, W, were well correlated across segments. There was 88% agreement on the existence of connections between segments, on average, and R = 0.72 correlation between the weights themselves.

To assess the accuracy of our model in predicting spikes, we also used goodness-of-fit tests based on the time-rescaling theorem [16]. In this test, the integral of the conditional intensity function over each inter-spike interval, z, should be drawn from a uniform distribution after rescaling (see [17] for more details). KS-tests for the best predicted neuron, worst predicted neuron, and a typical neuron are shown in figure (a). The sorted KS-statistics (the supremum of the point-wise differences between the CDF of z and the CDF for the uniform distribution) for the entire ensemble are shown in (b).
A more general model

\[ \lambda_i(t) = \exp \left( \sum_{i=1}^{N} \sum_{\Delta t=1}^{T} W_{i,\Delta t} s_i(t - \Delta t) + \sum_{k=1}^{K} V_{i,\Delta t} v_k(t - \Delta t) \right) \]

- Neurons are affected by other neurons
- Also affected via tuning curves by outside world
Graphical version

[Diagram showing a process involving movement/stimuli (external covariate), coupling with other neurons, nonlinearity, and predicted spiking, with spike counts and trials indicated.]
Information per Neuron

![Graph showing information per neuron for different network sizes and cortex types: Motor Cortex and Visual Cortex. The x-axis represents network size on a log scale, and the y-axis represents cross-validated log-likelihood (bits per neuron). The graph includes blue and red lines indicating receptive/movement field and interactions, respectively.](image)

Hatsopoulos, O(n log n), Kohn
Spike Timing Dependent Plasticity

![Distribution of multiplicative strength with respect to pre-to-post spike timing difference. A total of 372 pairs were analyzed.](image)
Tuning curves are explained away
Countless future machine learning problems

- Find structure in generative model
- Meaningful priors
- Link to cognitive phenomena
- Reverse engineer learning rules
Acknowledgements

- Ian Stevenson
- Nicho Hatsopoulos
- Adam Kohn
- Lee Miller, Jim Rebesco
- Sara Solla