Learning Predictive Filters

Lane McIntosh
Neurosciences Program, Stanford University School of Medicine

Introduction

- How would a system intent on only keeping information maximally predictive of the future filter its input?
- We analytically derive the optimal predictive filters for Gaussian stimuli by modifying [1] and [2]
- We learn the optimal predictive filters for non-Gaussian naturalistic movies
- We examine the role of prediction in the retina and compare our predictive filters to measured filters in salamandar retina

Machine learning interpretation:

We want to learn models with low complexity that generalize well to future data

Neuroscience interpretation:

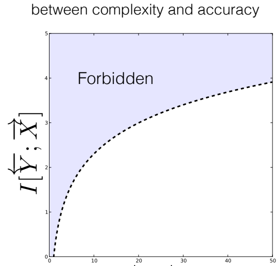
Neural circuits should allocate resources to encode bits that will be useful for the future

Predictive Information Bottleneck

Problem:

$$\min I[\overleftarrow{X};\overleftarrow{Y}] - \beta I[\overleftarrow{Y};\overrightarrow{X}]$$

Beta parameterizes the tradeoff between complexity and accuracy

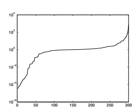


System:

$$x_{t+1} = Ax_t + \eta$$
$$y_{t+1} = C_\beta x_{t+1} + D_\beta y_t + \xi$$

We learn the linear filters C_β and D_β as a function of how many bits we can keep

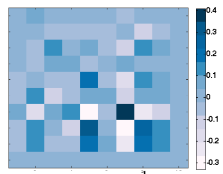
$$\overleftarrow{X} = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-k+1} \end{bmatrix} \quad \overrightarrow{X} = \begin{bmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+k} \end{bmatrix}$$



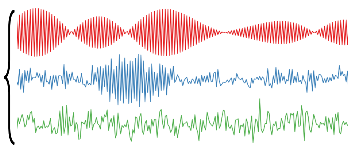
$$\overleftarrow{Y} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-k+1} \end{bmatrix}$$

Analytical Results

$$\min_{H(\beta)} (1-\beta)\log\left|H(\beta)\Sigma_{\overleftarrow{X}}H(\beta)^T + I\right| + \beta\log\left|H(\beta)\Sigma_{\overleftarrow{X}|\overrightarrow{X}}H(\beta)^T + I\right| \qquad \overleftarrow{Y} = H(\beta)\overleftarrow{X}$$



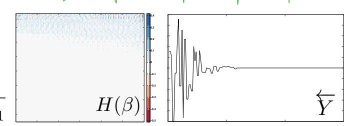
$$H(\beta) = \frac{1}{k}\begin{bmatrix} \sum_{i=0}^{t-1} D^i_\beta C_\beta A^{t-i} \\ \vdots \\ \sum_{i=0}^{t-k} D^i_\beta C_\beta A^{t-i} \end{bmatrix}\left[(A^{t-1})^{-1} \dots (A^{t-k})^{-1}\right]$$



$$H^*(\beta) = \begin{bmatrix} \alpha_1 v_1^T \\ \vdots \\ \alpha_k v_k^T \\ \vec{0} \end{bmatrix} \quad \alpha_i = \sqrt{\frac{\beta(1-\lambda_i)-1}{\lambda_i v_i^T \Sigma_{\overleftarrow{X}} v_i}}$$

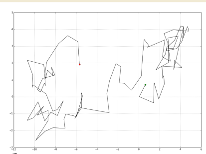
where

$$\frac{1}{1-\lambda_k} \leq \beta \leq \frac{1}{1-\lambda_{k+1}}$$

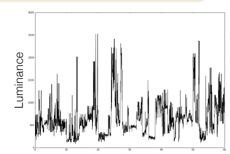


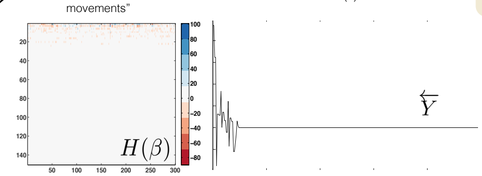
Learning predictive filters for naturalistic movies

Series of "saccades"

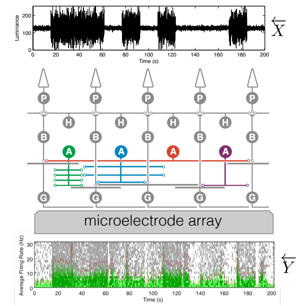


Interleaved with fixational "eye movements"

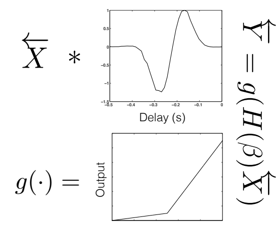




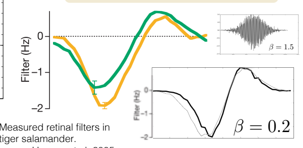
Predictive Filters in the Retina



Linear-Nonlinear model

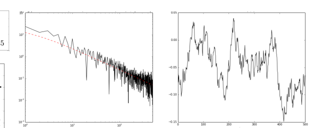


Real vs. Optimal filters

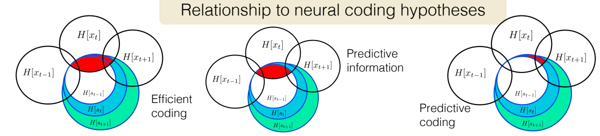


Measured retinal filters in tiger salamander. Hosoya et al. 2005

Stimulus



Relationship to neural coding hypotheses



Conclusions

- Neural systems and machine learning models have common goals of being generalizable with low complexity.
- The predictive information bottleneck provides a model-free way of quantifying this objective.
- The critical values of β provide an expectation for the different information-processing regimes we expect to find natural systems.

1. Chechik et al., 2005 "Information Bottleneck for Gaussian Variables." *Journal of Machine Learning Research*
2. Creutzig et al., 2009 "Past-future Information Bottleneck for Dynamical Systems." *Phys. Rev. E* 79, 041925