Nonlinear receptive field analysis: making kernel methods interpretable

Wolf Kienzle, Jakob H. Macke, Felix A. Wichmann, Bernhard Schölkopf, Matthias O. Franz

Max Planck Institute for Biological Cybernetics

Identification of stimulus-response functions is a central problem in systems neuroscience and related areas. Prominent examples are the estimation of receptive fields and classification images [1]. In most cases, the relationship between a high-dimensional input and the system output is modeled by a linear (first-order) or quadratic (second-order) model. Models with third or higher order dependencies are seldom used, since both parameter estimation and model interpretation can become very difficult.

Recently, Wu and Gallant [3] proposed the use of kernel methods, which have become a standard tool in machine learning during the past decade [2]. Kernel methods can capture relationships of any order, while solving the parameter estmation problem efficiently. In short, the stimuli are mapped into a high-dimensional feature space, where a standard linear method, such as linear regression or Fisher discriminant, is applied. The kernel function allows for doing this implicitly, with all computations carried out in stimulus space. As a consequence, the resulting model is nonlinear, but many desirable properties of linear methods are retained. For example, the estimation problem has no local minima, which is in contrast to other nonlinear approaches, such as neural networks [4].

Unfortunately, although kernel methods excel at modeling complex functions, the question of how to interpret the resulting models remains. In particular, it is not clear how receptive fields should be defined in this context, or how they can be visualized. To remedy this, we propose the following definition: noting that the model is linear in feature space, we define a nonlinear receptive field as a stimulus whose image in feature space maximizes the dot-product with the learned model. This can be seen as a generalization of the receptive field of a linear filter: if the feature map is the identity, the kernel method becomes linear, and our receptive field definition coincides with that of a linear filter. If it is nonlinear, we numerically invert the feature space mapping to recover the receptive field in stimulus space.

Experimental results show that receptive fields of simulated visual neurons, using natural stimuli, are correctly identified. Moreover, we use this technique to compute nonlinear receptive fields of the human fixation mechanism during free-viewing of natural images.

References

[1] J. D. Victor, Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy, Nature Neuroscience, 2005

[2] B. Schölkopf, A. J. Smola, Learning with kernels, MIT Press, 2002

[3] M. C. K. Wu, J. L. Gallant. What's beyond second order? Kernel regression techniques for nonlinear functional characterization of visual neurons, Society for Neuroscience, 2004

[4] R. Prenger, M. C. K. Wu, S. V. David, J. L. Gallant, Nonlinear V1 responses to natural scenes revealed by neural network analysis, Neural Networks, 2004