Score Matching

Estimation of a Two-Layer Model of Natural Images

Urs Köster, Aapo Hyvärinen and Silke Dreissigacker

Neuroinformatics Group, Basic Research Unit, Helsinki Institute for Information Technonlogy, University of Helsinki, Finland

The Scope

Previous work on extending *ICA models* with a second layer has usually been done with a fixed second layer, e.g. pooling, for simplicity (see [1] and [2]). Here we present a two-layer ICA model that uses the novel technique of Score Matching [3] for *estimating both layers simultaneously*. This allows for the first time to obtain *complex cell responses* from an ICA model without making the pooling explicit. The results indicate that a continuum ranging from simple to complex cells is obtained.

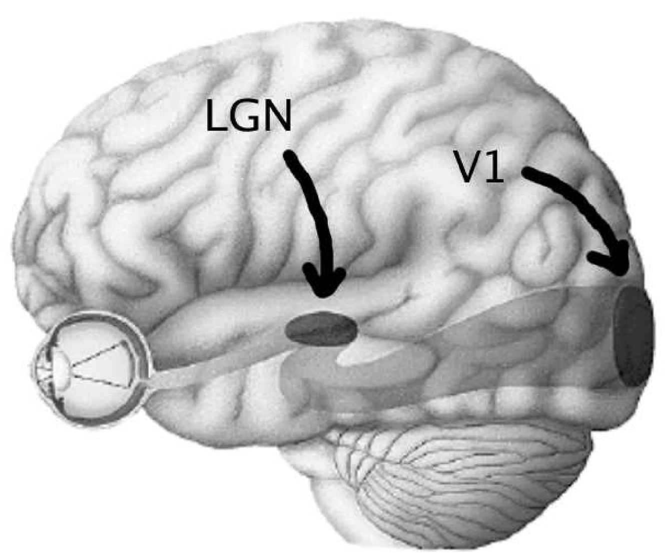

FIGURE 1: The location of **primary visual cortex** in the human brain.

- We developed a new, powerful method, Score Matching, for the estimation of multi-layer unsupervised models.
- We present a two-layer model where both layers can be estimated simultaneously.
- Applying the model to **natural images** allows us to model processing in **primary visual cortex**.
- Score Matching allows the **estimation of non-normalized *energy-based* models** consistently without the need for approximations or Monte Carlo methods.

Score Matching

For a general probability distribution

the intractable normalization constant Z. The model is consistently estimated by minimizing the *distance between the model and data score function*. The squared distance between the *model score function* $\Psi(\eta;\Theta)$ and the *data score function* $\Psi_{\mathbf{x}}(.)$ is given by

$$J(\Theta) = \frac{1}{2}\int_{\eta\in\mathbb{R}^n} p_{\mathbf{x}}(\eta)\|\Psi(\eta;\Theta) - \Psi_{\mathbf{x}}(\eta)\|^2 d\eta \quad (3)$$

This is not practical, because estimating the data score function is a difficult nonparametric problem. Using partial integration, the distance between the model and data score functions can be computed solely in terms of derivatives of the model score function. This is easy to compute.

$$\tilde{J}(\Theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[\frac{\partial\psi}{\partial\mathbf{x}} + \frac{1}{2}\Psi_i^2(\mathbf{x}(t);\Theta)\right] + C \quad (4)$$

This allows the optimization to be performed by virtually any gradient method.

Two layer model

To demonstrate the power of Score Matching we apply it to a **general two layer neural network model** of the form

$$\log p(x|W,V) = \sum_{h=1}^{m}\sigma\left[V_h g\left(W\mathbf{x}\right)\right] \quad (5)$$

where W and V are weight matrices, $g()$ and $\sigma()$ scalar nonlinear functions and the index h runs over the second layer filters.

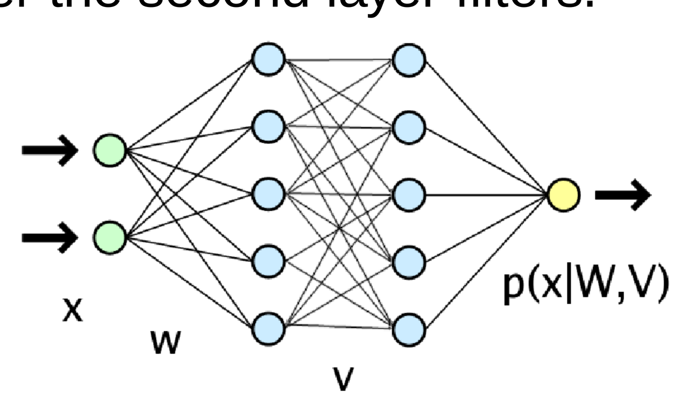

FIGURE 2: Topology of the two-layer neural network.

it is important to choose these with sufficient smoothness.

- For the experiments here, the first nonlinearity was $g(\mathbf{x}) = \mathbf{x}^2$ and the second $\sigma(\mathbf{x}) = \sqrt{\mathbf{x}+1}$. This is similar to the nonlinearities used in our previous work on ISA-models [2].

Artificial data

As a first test for the estimation method we have generated data according to the **ISA model**. This is supergaussian data with dependencies within, but not between subspaces of the data variables. This data was then mixed with a random mixing matrix A. We used 10.000 samples of 21-dimensional data generated with a subspace size of three.

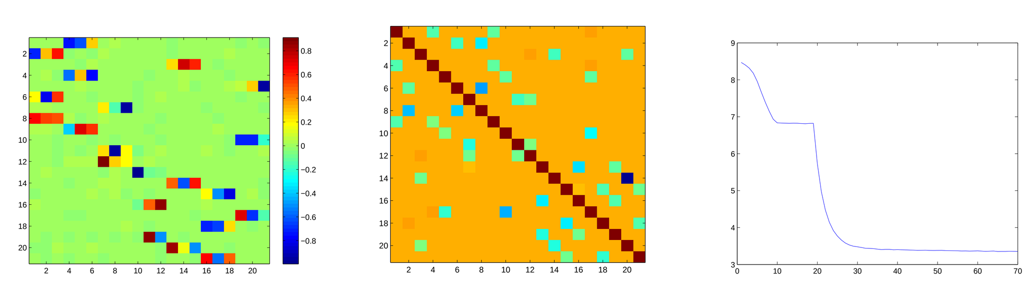

FIGURE 3: The model was tested with ISA data, convergence is fast and finds a good local minimum. We show (a) the product of mixing and demixing matrix, (b) the second layer weights and (c) the convergence plot of fist W and then V.

Figure two shows how the first layer weights W invert the mixing up to subspace membership, while V determines which variables belong together in one subspace.

Natural image data

- We use 20.000 image patches of 12×12 pixels to learn the statistical structure of natural image data.
- The data was whitened and *contrast gain*

learns connections between similar first layer features.

- **Complex cell type receptive fields are obtained from static natural images** without pre-supposing the pooling.

Results

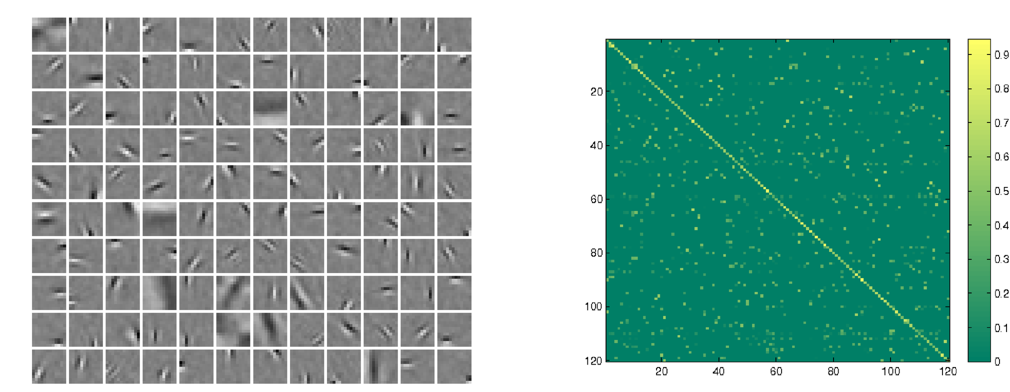

FIGURE 4: First layer filters show the classical Simple-Cell type structure. Connection in the second layer are sparse, with connections between similar units.

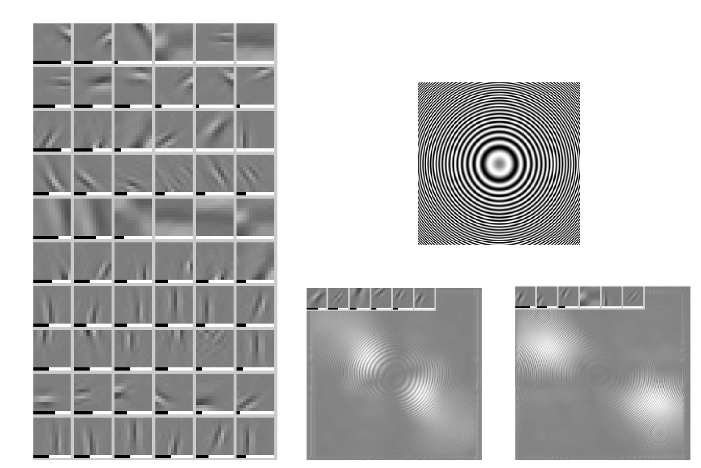

FIGURE 5: (a) The connectivity of randomly selected complex cells is show. The black bar indicates the contribution of a particular filter to the cell.
(b) Two cells are analyzed by scanning them over a circular grating

Future prospects

- The method can easily be extended to **three and more layers**.
- Gearing the **nonlinearities** towards specific outputs; using **adaptive nonlinearities**.
- Applying the to a variety of data other than