Introduction

Neural activity encodes multiple-cause stimuli with discrete events. Neurons either spike or remain inactive. Many modeling approaches therefore rely on binary units for encoding. Prominent examples are, for instance, restricted Boltzmann machines [5] and more recently, deep belief networks [6]. In this work we study a probabilistic generative model with binary units. We investigate the component extraction capabilities of a model with hidden and observed layer both encoding binary data through Bernoulli distributions. In this setting basis functions cannot be combined using summation as in sparse coding models [3] but require non-linear combination rules.

BMCA: Binary Maximal Causes Analysis

In this work we explore a generative model with the combination of binary variables using a maximum combination (BMCA), previously used for continuous variables, (1,2), and optimize the model parameters using EM. As in standard approaches such as Sparse Coding [3], Independent Component Analysis, or Binary Sparse Coding (BSC), MCA assumes a sparse prior with independent hidden variables.

\[ p(\vec{s} | \Theta) = \prod_{h=1}^{H} \pi_h \{ 1 - \pi_h \}^{1 - s_h} \]

\[ p(\vec{y} | \vec{s}, \Theta) = \prod_{d=1}^{D} W_d(\vec{s})^{Y_d}(1 - W_d(\vec{s}))^{(1 - Y_d)} \]

\[ W_d(\vec{s}) = \max_h \{ s_h W_{dh} \} = \lim_{\rho \to \infty} \left( \sum_h (W_{dh} s_h)^\rho \right)^{\frac{1}{\rho}} \]

\[ \vec{y} \in \{ 0, 1 \}^D \] observed variables

\[ \vec{s} \in \{ 0, 1 \}^H \] hidden variables

\[ W \in [0,1]^{D \times H} \] basis functions

with M-step update equations for basis functions \( W \):

\[ W_{dh}^{\text{new}} = \frac{\sum_n (C_{dh}) Y_d^{(n)}}{\sum_n (C_{dh})} \]

\[ C_{dh} = \frac{s_h}{W_d(\vec{s}) - W_d(\vec{s})^2} \left( \frac{W_{dh}}{W_d(\vec{s})} \right)^{\rho - 1} \]

In the place where standard sparse coding approaches, NMF, or ICA use the sum to combine basis functions, BMCA uses a (pixel-wise) maximum operation. To derive tractable approximations for parameter estimation we apply Expectation Truncation (ET; [4]) - a variational EM approach which reduces the hidden space to only those variables contributing most posterior mass. This allows one to infer all model parameters, namely the basis functions \( W \) and the degree of sparseness, \( \pi \). The resulting algorithm is applicable to large-scale problems with hundreds of observed and hidden variables.

Application to MNIST digits database

The non-linearity of the BMCA is illustrated above using an extracted patch from a natural scene; shown is the contrast between the non-linear max of BMCA (\( \max_h \{ s_h W_{dh} \} \)) and the sum of standard superposition (\( \sum_h (s_h W_{dh}) \)). This may represent a more plausible assumption for the superposition of hand-strokes in the construction of digits.

To study the implications of the non-linear superposition for visual data, the BMCA algorithm was applied to \( N = 40,000 \) MNIST digits with \( D = 28 \times 28 = 784 \) pixels, \( H = 64 \) hidden units, and \( H' = 8 \).

Inferred basis functions (H=64):

Conclusions

- This work explored an approach that combines a fast preselection of relevant data components with a subsequent recurrent processing phase (compare [7]), and has recently been linked to neural processing.
- Variational training in the model infers the weights of the connections between the hidden and the observed units as well as the prior activation probabilities of the hidden units.
- In numerical experiments on artificial data and more realistic data, we show that components of mixtures in binary data can successfully be recovered and that such experiments can be scaled to high dimensional observed and hidden spaces.
- Future work will explore different non-linear combinations of basis functions with hidden units, like that used in RBMs.

References