Efficient Learning and Inference of Sparse Overcomplete Representations

Koray Kavukcuoglu[1], **Marc'Aurelio Ranzato**[1], **Y-Lan Boureau**[1,2]
Yann LeCun[1]

[1] The Courant Institute of Mathematical Sciences - New York University
715 Broadway, 12th Floor
New York, NY 10003
[2] INRIA Rocquencourt, WILLOW team
koray@cs.nyu.edu, ranzato@cs.nyu.edu, ylan@cs.nyu.edu, yann@cs.nyu.edu

Sparse overcomplete representations are useful in many vision applications, such as feature extraction [7], recognition [6], denoising, and inpainting of natural images [2]. Many such algorithms have focused on learning a dictionary of basis functions such that any image patch can be reconstructed as a linear combination of a small subset of them. Unfortunately, finding the coefficients corresponding to a new image patch is expensive because it involves optimizing the reconstruction error under non-quadratic sparsity constraints (such as an L1 penalty) [5, 3, 4].

We propose an algorithm that learns basis functions in such a way that the optimal set of coefficients representing the input patch can be efficiently computed by a feed-forward pass through a simple parametrized function (an encoder), without requiring any optimization.

The coefficient vector $z \in \mathbb{R}^n$ is used to model the input space $y \in \mathbb{R}^m$ in terms of a basis set $\Phi \in \mathbb{R}^{m \times n}$. We use ℓ^2 norm for penalizing the reconstruction error, and ℓ^1 norm for enforcing the sparsity of the coefficient vector:

$$\|y - \Phi z\|_2^2 + \alpha_z \|z\|_1$$

which is equivalent to the solution using approximate decomposition of input signals as given in Basis Pursuit algorithm [1]. Based on this formulation we propose to learn the basis set and an encoder for efficient inference of the coefficient vector by minimizing over the following loss functional:

$$\mathcal{L}(y, z, \Phi, \Psi) = \|y - \Phi z\|_2^2 + \alpha_z \|z\|_1 + \alpha_e \|z - \Psi(y)\|_2^2$$

where $\Psi(y) : \mathbb{R}^m \mapsto \mathbb{R}^n$ is the encoder function, and the columns of Φ are *constrained* to have unit ℓ^2 norm. The first term in the above loss penalizes the reconstruction error, the second term penalizes the sparsity of the representation, and the last term ensures that coefficient vector is predictable by the encoder.

We present an in-depth experimental comparison between machines trained with different encoders, e.g. linear and non-linear encoders. We also compare different training algorithms, such as standard back-propagation of the error and deterministic EM, which treats the code as latent variable in a block coordinate descent algorithm. The latter is implemented as on on-line learning algorithm as follows:
1. minimize $\mathcal{L}$ with respect to z, thus obtaining the coefficient vector z^* that gives best reconstruction under sparsity and predictability constraints.
2. update parameters of Ψ and Φ using the optimal code z^*.
3. normalize the columns of Φ to have unit ℓ^2 norm.

EM training, as explained above, yields basis functions that look like gabor filters and are similar to filters usually found with algorithms optimized for learning useful sparse features, while training with backpropagation leads to more localized basis functions that might not be useful for applications other than mere reconstruction.

We evaluate these different methods by using both the "unsupervised" objective that was used to train them, as well as a "supervised" objective, such as recognition error of a classifier trained on the learned representation. Results depend on how tough the learning task is. If the code is not highly overcomplete, even a linear encoder trained by back-propagation of the error can give acceptable accuracy.

We demonstrate our algorithm by learning highly overcomplete dictionaries of basis functions from a set of natural image patches as shown for instance in fig. 1, and by applying the algorithm as building block for training a deep network that efficiently recognizes objects in natural images.

References

[1] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

[2] Michael Elad and Michal Aharon. Image denoising via learned dictionaries and sparse representation. In *CVPR '06*, pages 895–900, Washington, DC, USA, 2006. IEEE Computer Society.

[3] Honglak Lee, Alexis Battle, Raina Rajat, and Andrew Y. Ng. Efficient sparse coding algorithms. In *NIPS 19*, 2007.

[4] Joseph F. Murray and Kenneth Kreutz-Delgado. Learning sparse overcomplete codes for images. *The Journal of VLSI Signal Processing*, 45:97–110, 2008.

[5] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision Res*, 37(23):3311–3325, 1997.

[6] Marc'Aurelio Ranzato, Fu-Jie Huang, Y-Lan Boureau, and Yann LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *CVPR '07*. IEEE Press, 2007.

[7] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann LeCun. Efficient learning of sparse representations with an energy-based model. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *NIPS 19*. MIT Press, Cambridge, MA, 2007.

Topic: learning algorithms, unsupervised learning
Preference: oral

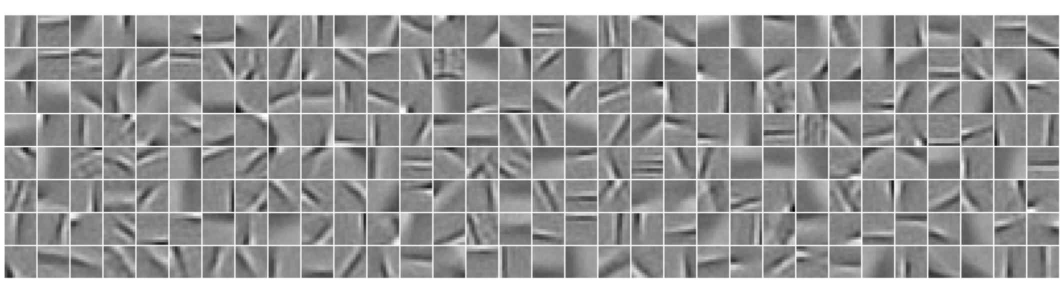

Figure 1: 256 basis functions learned by deterministic EM algorithm trained on 12x12 natural image patches randomly extracted from the Berkeley dataset.