Progress Report
16-28 August 2011

Raj K. Kumar, Amir Hesam Salavati
E-mail: raj.kumar@epfl.ch, hesam.salavati@epfl.ch

Supervisor: Prof. Amin Shokrollahi
E-mail: amin.shokrollahi@epfl.ch

Algorithmics Laboratory (ALGO)
Ecole Polytechnique Federale de Lausanne (EPFL)

August 29, 2011

1 Summary

During the last week, we were mostly focused on reading some new papers that adress ideas related to learning the constraint matrix of a constraint satisfaction problem from the patterns that satisfy those constraints. Although the following papers do not address the exact problem that we are interested in, with slight modifications they can be suitable for our purpose.

2 The subspace learning problem

Continuing the same path on reading more about the subspace learning problem, i.e. the problem of classifying patterns when they form a subspace, we read [1]. In [1] the authors propose two neural networks to solve the Dual Subspace Pattern Recognition (DSPR) problem. In this framework, some classes are represented by their basis vectors while the others are represented by their null space basis. Now the problem of interest is to identify these basis and null basis vectors from a set of given training data. In [4] the same problem was addressed but only for the case that all classes are represented by their own basis vectors and not those that are orthogonal to them.

The proposed method to solve the DSPR problem are based on the Stochastic Gradient Ascent (SGA) due to Oja, [5] and the Asymmetric Lateral Inhibition Algorithm (ALIA) by Rubner et al. [6]. Both of the approaches use simple techniques to find a basis vector in the subspace and then find the projection of the remaining patterns over the already found set of basis vectors in order to find the rest of basis set. More specifically, the first one, employs a network that has K subnets, one for each class. The input is given to each subnet and the outputs of the subnets go to a winner-take-all unit which specifies which class the given query belongs to. Within subnet i, there are p_i units that specify the basis vectors of subspace i, where p_i is dimension of the subspace. Now in order to learn the subspaces corresponding to each class from the training set, two learning rules are suggested by the authors: one for the subspaces represented by their own basis and the other one for those defined by their dual (null) space. In words, these equations make sure that the first unit of subnet i will converge to an orthogonal direction of the input data x. The second unit will converge to the orthogonal direction of the *novel* part of x with respect to the first orthogonal component and so on.

The second suggested approach is basically the same as the first model except that the input to all the layers are the same this time, namely it is the training pattern. And in order to find the orthogonal set of basis vectors, there are some feedback signals that affect the next layers. The learning rules for the ALIA model are quite similar to their counterparts in the SGA model and it can be shown that after the end of learning process, the weight vectors are the same as those given by the SGA model. So both approaches yield the same results.

Using simulations, they have shown that the proposed approaches has a remarkable performance in identifying subspaces in the training set. Furthermore, they show that their approach is better than the conventional back propagation technique, which is slower, less reliable in classification and have a higher computational cost in this particular problem.

Principle Component Analysis (PCA) is another method to learn the basis vectors of a given set of patterns. In [3], the authors investigate the landscape of the quadratic cost functions for *linear* neural networks in PCA. More specifically, the problem of interest is to see how the energy landscape $E = \sum_t \|y^{(t)} - ABx^{(t)}\|^2$ looks like for a neural network with n input and output neurons as well as a hidden layer of p neurons. $x^{(t)}$ and $y^{(t)}$, $1 \leq t \leq T$, are the input and desired output patterns. The matrices $B_{p\times n}$ and $A_{n\times p}$ are the weight matrices from input to the hidden layer and from the hidden to the output layer, respectively.

The authors do not focus on a particular learning algorithm, like back propagation, but investigate how the energy landscape looks like in general. They show that this particular type of networks results in an energy landscape with a *unique global minimum*. This finding justfies the assumption of using back propagation to implement gradient descent for minimizing the error cost function since the way these algorithms are implemented, it is unlikely to get stuck in saddle points (although theoretically it is possible).

3 Sparse Learning

For a while, we have been trying to find a way to enforce the sparsity constraint in the learning process. More specifically, there are algorithms to learn a set of contraints from the set of training patterns. However, they do not necessarily lead to sparse connectivity matrices, which is the case of interest for us. In [2] the author address a similar problem but in a different setting. The problem the authors are interested in is representing images with a set of basis functions that yield efficient compression. The efficiency is in the sense that much information in the original image is preserved with these bases while most of the coefficients that describe each image in terms of the bases are zero for any image. Furthermore, their approach results in bases that are similar to humans visual receptive fields.

The sparsity in information preservation is achieved by minimizing a cost function, given by equation (1), which minimizes information loss and the number of non-zero coefficients at the same time.

$$\min E = \min sum_{x,y} \left[I(x,y) - \sum_i a_i \phi_i(x,y) \right]^2 + \lambda \sum_i S(\frac{a_i}{\sigma}) \tag{1}$$

Where $S(.)$ is nonlinear function and σ is a scaling factor.

The goal is find the basis functions ϕ_i as well as the corresponding coefficients a_i. The learning is accomplished in two phases. For *each image* in the training phase, the following steps are performed:

1. Minimize the cost function (1) with respect to coefficients a_i's.
2. Use gradient descent to update proper bases.

Using simulations, the authors show the effectiveness of the proposed method as well as its similarity to human visual receptive fields.

4 Conclusions and future works

We have read some papers mostly about learning the null space basis of a set of given patterns using neural networks. There are a number of neural approaches to find the subspaces, like [1] and [4], but they do not necessarily yield sparse neural graphs. On the other hand, we have [2] which enforces sparsity during the learning phase for a different problem. We would like to work on ideas similar to the one mentioned in [2] in order to enforce sparsity in our setting.

References

[1] L. Xu, A. Krzyzak, E. Oja, *Neural nets for dual subspace pattern recognition method*, Int. J. Neur. Syst., Vol. 2, No. 3, 1991, pp. 169-184.

[2] B. A. Olshausen, D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, Nature Vol. 381, 1996, pp. 607-609.

[3] P. Baldi, K. Hornik, *Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima*, Neural Networks, Vol. 2, 1989, pp. 53-58.

[4] E. Oja, T. Kohonen, *The subspace learning algorithm as a formalism for pattern recognition and neural networks*, Neural Networks, Vol. 1, 1988, pp. 277-284.

[5] E. Oja, *Principal components, minor components, and linear neural networks*, Neur. Net., Vol. 5, No. 6, 1992, pp. 927-935.

[6] J. Rubner, P. Tavan, *A self-organizing network for principal-component analysis*, Europhysics Letters, Vol. 10, No. 7, 1989, pp. 693-698.