POLITECNICO DI TORINO

Dipartimento di Scienza Applicata e Tecnologia
Master program in Physics of complex systems

Storage properties and hard-spheres packing in Perceptron and Support Vector Machines

Author: Giovanni Catania

Supervisor: Silvio Franz

Co-Supervisor: Riccardo Zecchina

Summary

The goal of this work is to provide a theoretical analysis of the storage properties of two different models, which are usually used as supervised learning algorithms: perceptron and support vector machines.
The perceptron is a very simple model of a single-layer neural network, introduced to describe the response of a neuron to a given set of stimuli. In a machine learning perspective instead, the perceptron allows to perform binary classification of data according to a linear rule. Historically, physicists' interest in neural networks arised mainly from the analogy with simple magnetic disordered systems, such as spin-glasses. From a theoretical viewpoint, one of the most pioneering works about the statistical properties of the perceptron was provided by E. Gardner: it was pointed out that the storage properties of a perceptron are intimately connected with the performance of attractor neural networks when used as associative memories. The approach developed is the same used to study the properties of spin glasses, whose main ingredient is the

well-known *replica method*. From that point, a lot of research has been carried out to study generalization and learning properties of perceptron.
More recently, a new way to interpret this model has been studied: indeed, starting from the fact that the storage problem for perceptron can be considered as a constraint satisfaction problem (CSP), it is possible to establish an equivalence with a simple model for jamming of hard-spherical object in an infinite dimensional space. The equivalence is valid only in a particular regime (i.e. when the stability parameter k is negative).

From a machine learning perspective instead, the most important limitation of this algorithm is due to its ability to classify only linearly separable data. Therefore, several attempts have been done to improve its performances. Support vector machines have been the most important improvement in the framework of pattern classification and supervised learning. In addition to linear classification, SVMs can efficiently perform a non-linear classification of data by mapping their inputs into higher-dimensional feature spaces.
This work provides an analytical study of the storage problem in both models, through the statistical mechanics approach mentioned above.
The work and the corresponding report are divided in two main sections.

SECTION I
The first part of the work focuses on a generalized version of the standard perceptron model. In particular, the *stability parameter* k is considered as a random variable rather than a constant, in the same way as the input components. Three cases have been considered for what concerns the probability distribution of such parameter: a gaussian distribution, a continuos uniform distribution and a discrete, m-valued distribution. The starting point of the analysis is the computation of the *Gardner volume* corresponding to the satisfied constraints. The statistical mechanics approach requires the evaluation of this quantity at thermodynamical limit, which is obtained when the number of degrees of freedom goes to infinity. For each of the three cases defined above, the the calculations have been exploited by means of replica method, working always in replica symmetry (RS) assumption. At the end, the following quantities have been computed as functions of the parameters related to the specific probability distribution for k: quenched entropy and critical storage capacity. The saddle point equations for the order parameter have been numerically solved. Moreover, the stability of RS saddle point has been checked, using the standard method developed by Almeida and Thouless for the SK solution of a spin glass model. The results can be interpreted in both regimes in which the model describes the storage problem or the hard-spheres packing one. The most interesting case is the one

corresponding to the uniform distribution for k, because it describes an ensemble of spherical objects with different radius, generalizing the results already known in literature. These results can be used as a starting point to develop a deeper analysis of the model, for instance by using a replica symmetry breaking scheme. Finally, as a further development, simulations of hard spheres in high dimentional space could be carried out to check the validity of the model and verify again the equivalence between the two pictures.

SECTION II

The final part of the work is a starting point to generalize the storage problem when dealing with higher-order classification rules, which is the case of support vector machines. Starting from the storage problem in SVM I analyzed some possible source of errors in the results already present in literature, in particular in the case of a "quadratic" perceptron. Apparently, there are some "loop" corrections which are not negligible at thermodynamical limit. Of course, this is still an on-going work. However, in relation to the first part of the work, also in this models we expect to find regimes in which ergodicity breaks down and replica symmetry breaking ansatz are required. Therefore, the ultimate goal would be to establish a new equivalence between the storage problem in higher-order classification models and some kind of hard-sphere packing problem with "non-linear" spherical objects.

This work has been carried out at the Laboratoire de Physique Théorique et Modéles Statistiques (CNRS unit at Universite Paris-Sud – Orsay, France) under the supervision of professor Silvio Franz.