How to generate data for approximating multidimensional surfaces ? Application to the approximation of viability kernels

Benoît Gandar (1,2) & Gaëlle Loosli (2) & Guillaume Deffuant (1)

(1) Cemagref de Clermont-Ferrand
LISC (Laboratoire d'Ingénierie pour les Systèmes Complexes)
24, avenue des Landais - BP 50 085 - 63 172 Aubière Cedex 1 - France
firstname.surname@cemagref.fr - http ://wwwlisc.clermont.cemagref.fr

(2) LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes)
Complexe scientifique des Cézeaux, 63 173 Aubière cedex, France
gaelle@loosli.fr

This paper deals with active learning classification problems. This type of problems may arise in various contexts, but our main motivation is to compute policies of actions on a dynamic system, in order to keep the system viable. This type of problem is frequent in ecology, economics or robotics. From viability theory, we know that an essential step to solve these problems is to determine a particular subset of the state space called the viability kernel. An algorithm [2], computes viability kernels by recursive approximations using statistical learning algorithm (Support Vector Machines being a particularly relevant learning technique in this context). At each step of the algorithm, one must solve an active classification learning problem : we can compute if any point belongs to the set to approximate (i.e. we have an oracle), but this calculation can be very expensive, especially in high dimension problems. It is thus often essential to choose carefully the distribution of the learning examples, in order to maximize the quality of learning for a given size of the sample.

To solve this problem, we first tried to adapt the theoretical approach developed on the apparently very similar problem of active function learning (regression). This approach shows that low discrepancy samples are theoretically more adequate to maximise the quality of learning. Yet, we show that, surprisingly, these results cannot be transfered to classification. This is the first contribution of this paper. The second contribution is to show, with theorical and experimental arguments, that minimising the dispersion of sample is the relevant strategy to insure the best results in active classification learning problems.

1 The results about active regression learning do not apply to active classification learning

Using results about estimation of multidimensionnal integrals, Cervellera & Muselli show in [1] that generalisation error bounds are optimal using low discrepancy sequences as sample [1]. These bounds are based on the theorem of Koksma-Hlawka. Denoting f the target function (supposed to be real), $\hat{f}$ the approximation of f obtained by a learning algorithm (for instance through an empirical risk minimisation algorithm) on a learning set X, and $D^*(X)$ the discrepancy of X, the theorem establishes :

$$Learning\ Error = \int \left|\tilde{f} - f\right| \leq V_{HK}\left(\left|\tilde{f} - f\right|\right) D^*(X)$$

[1] Discrepancy of a sequence is, supposing that the sequence is defined in $[0,1]^d$ and considering any convex subset V of $[0,1]^d$, the largest difference between the volume of V and the proportion of elements of the sequence located inside V (see for instance [6] for details)

where $V_{HK}(g)$ is the variation of g in the sense of Hardy Krause : a measure of its regulartity. In this inequality, it is clear that error is determinism and directly proportional to the discrepancy of the learning set, supposing the variation is finite. The hypothesis of finite variation, which is very heavy, was studied and weakened by Mary in the case of regression (show [5]).

We show that this approach does not apply to the problem of active classification learning (learning manifold boundaries). Indeed, the superior bound in the previous inequality is equal to infinity, beacause of the infinite variation of indicator functions (shows [7]). This result is somehow suprising, because classification can be seen as a particular case of function approximation (regression) problem.

2 Low dispersion is a better criterion of sample quality for active classification learning

Lacking a satisfactory theorical approach, the relevance of low discrepancy sequences in classification can be assessed experimentally. Iwata & Ishii (in [4]) observed experimently a gain of quality in classification with the multi-layer perceptron using low discrepancy sequences rather than random sequences as learning sets. However, using SVMs, we observed in [3] a slightly inferior generalisation quality when learning on low discrepancy sequences than when learning on grids of the same size.

All this leads us to suspect that discrepancy is not the relevant property to consider for building optimal samples in active classification learning. In the particular case of a basic learning algorithm (variant of k earest neighbours), we prove theorically the generalisation error is lower than a linear function of the sample dispersion (the radius of the largest ball of $[0,1]^d$ not containing any point of the learning set). We conjecture that this property can be extended to other learning algorithms, hence that minimising dispersion is generally a good strategy for optimising active classification learning. We report a set of experiments on artificial forms of different regularity and in various dimensions, using SVMs as a learning algorithm, which support this conjecture.

Références

[1] C. Cervellera and M. Muselli. Deterministic design for neural network learning : An approach based on discrepancy. In *Proceedings IEEE Transactions on neural network*, volume 15, pages 533–544, 2004.

[2] G. Deffuant, L. Chapel, and S. Martin. Approximating viability kernels with support vector machines. In *In IEEE transactions on automatic control*, 2007.

[3] B. Gandar, G. Deffuant, and G. Loosli. Les suites à discrépance faible : un moyen de réduire le nombre de vecteurs supports des svms ? In 12^{ieme} *Journée Scientifique de l'Ecole Doctorale SPI : Apprentissage statistique - Apprentissage symbolique. Annales scientifiques de l'Université Blaise Pascal, Clermont-Ferrand II.*

[4] K. Iwata and N. Ishii. Discrepancy as a quality measure for advoiding classification bias. In *Proceedings of the 2002 IEEE International Symposium on Intelligent Control. Vancouver, Canada.*, volume 15, pages 533–544, 2002.

[5] J. Mary. *Etude de l'Apprentissage Actif, Application à la Conduite d'Expériences*. PhD thesis, UniversitéParis XI, 2005.

[6] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Ed. Society for Industrial and Applied Mathematics, 1992.

[7] Owen. Multidimensional variation for quasi-monte carlo. 2004.