Journal de la Société Française de Statistique
Vol. 160 No. 3 (2019)

Complex penalties and slope heuristics

Titre: Pénalités complexes et heuristique de pente

Bertrand Michel[1]

Since the seminal works of Birgé and Massart (2001, 2007), important progress have been made to understand and generalize the principles of the slope heuristics method. Unfortunately, less effort have been made on the practical aspects of the method. This new survey by Sylvain Arlot is a very important contribution to the field, not only because it gives a complete and comprehensive presentation of what can be proved on the slope heuristics, but also because it clarifies the different versions of the slope heuristics that can be implemented in practice.

Penalty shapes used with the slope heuristics typically correspond to the number of parameters. However, in many cases the exact expression of penalty shapes provided by model selection theorems are more complex, that makes them difficult to be used directly with the slope heuristics. It is for instance the case when deriving penalty functions from entropy calculations. I give below a few illustrations from my own experience of the slope heuristics for which complex penalty shapes occur.

Bracketing entropy calculations for Gaussian mixtures models produce penalty shapes that combine the parameters in a non trivial way (Maugis and Michel, 2011). One simple option in this situation is to upper bound all the penalty terms by a dominating term proportional to the total dimension. In the context of model selection for geometry inference (Caillerie and Michel, 2011) and geometry data analysis (Chazal et al., 2016), many unknown geometric quantities appear in the penalties. Simplifying the penalty terms for applying the slope heuristics is also necessary in this framework. Another example of complex penalty is for linear regression with Gaussian stationary error process (Caron, 2019). In this context, the penalty term involves a truncated trace of the covariance matrix of the errors, which is of course unknown in practice.

These examples illustrate the fact that, in general, we have to simplify penalty shapes to be in position of applying the slope heuristics. It is not clear if simplifying the penalty by this way may provide a measure a complexity which is appropriate for applying the slope heuristics. Hopefully, from a practical point of view, an important advantage of the slope heuristics is that it is very easy to detect when the method goes wrong. When there is no clear jump and no clear linear behavior with respect to chosen complexity, it suggests to reconsider the penalty shape. In this perspective it would be very useful to provide well-founded and data-driven methods for the validation of penalty shapes.

[1] École Centrale Nantes

References

Birgé, L. and Massart, P. (2001). A generalized Cp criterion for Gaussian model selection. *Technical report, Universités de Paris 6 et Paris 7*, Prépublication 647.

Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probability Theory and Related Fields*, 138(1-2):33–73.

Caillerie, C. and Michel, B. (2011). Model selection for simplicial approximation. *Foundations of Computational Mathematics*, 11(6):707–731.

Caron, E. (to appear, 2019). *Comportement des estimateurs des moindres carrés du modèle linéaire dans un contexte dépendant. Étude asymptotique, implémentation, exemples*. PhD thesis, École Centrale de Nantes.

Chazal, F., Giulini, I., and Michel, B. (2016). Data driven estimation of Laplace-Beltrami operator. In *Advances in Neural Information Processing Systems*, pages 3963–3971.

Maugis, C. and Michel, B. (2011). Data-driven penalty calibration: a case study for Gaussian mixture model selection. *ESAIM: Probability and Statistics*, 15:320–339.