Abstracts

V-fold penalization: an alternative to V-fold cross-validation

Sylvain Arlot

One of the most widely used model selection techniques is *V-fold cross-validation* (Geisser [Gei75]). It estimates the prediction error of estimators built upon $n(V-1)V^{-1} < n$ data, which can be interpretated as overpenalization. From the asymptotical viewpoint, this can be suboptimal (when V is fixed) and it has to be corrected, for instance following Burman [Bur89]. However, when the sample size is small, it may happen that $V = 2$ gives better results than $V = 10$, because overpenalization is benefic in some cases [Arl07a]. The choice of V in V-fold cross-validation can then be a difficult problem.

Following Efron's resampling heuristics [Efr79], we propose to use a V-fold resampling scheme to define a new penalization procedure, called *V-fold penalization* ([Arl07b], Chap. 5). It generalizes Burman's bias correction, and produces a flexible procedure, where V is decoupled from the overpenalization factor.

In the framework of regression on a random design with heteroscedastic noise, we prove several non-asymptotic results about V-fold subsampling, and more general resampling schemes. In particular, V-fold penalization (with V fixed) satisfies a non-asymptotic oracle inequality with constant almost one, which implies its asymptotic optimality. Hence, it improves on V-fold cross-validation. Moreover, choosing a particular family of models, we obtain an estimator adaptive to the smoothness of the regression function and the heteroscedasticity of the noise. Thus, V-fold penalties are more robust that Mallows' C_p criterion.

The theoretical results concerning V-fold penalties stay valid for resampling penalties with general exchangeable weights ([Arl07b], Chap. 6). In particular, they can be applied to V-fold penalties with $V = n$, as well as bootstrap penalties (defined by Efron [Efr83]). This extends an asymptotical result on bootstrap penalties in another framework (Shibata [Shi97]). Using independent Rademacher weights, one obtain a localized version of Rademacher complexities (Koltchinskii [Kol01] ; Bartlett, Boucheron and Lugosi [BBL02]) that is much easier to compute than local Rademacher complexities (Lugosi and Wegkamp [LW04] ; Koltchinskii [Kol06]).

Although we have to assume a particular structure for the models (*i.e.* they are all made of histograms), we believe that the same results hold in a much more general framework. We for instance have partial results for general bounded regression and binary classification ([Arl07b], Chap. 7).

A simulation study shows that V-fold penalties behave quite well in several cases. Moreover, they often outperform V-fold cross-validation and Mallows' C_p penalties, in particular in difficult heteroscedastic situations. Their flexibility allows to improve performances when the signal-to-noise ratio is small; this is obtained by taking V large enough, together with overpenalization.

The choice of V also appear to be quite easier: the performances of V-fold penalties are always better when V increases. Then, V has only to be chosen according to the computational complexity of the procedure, which is exactly the same as the one of V-fold cross-validation.

References

[Arl07a] Sylvain Arlot. Model selection by resampling penalization. Short version. Arxiv:math.ST/0701542, January 2007.

[Arl07b] Sylvain Arlot. *Resampling and Model Selection*. PhD thesis, University Paris XI, December 2007. Available at `http://www.math.u-psud.fr/~arlot/publi.html.en`

[BBL02] Peter L. Bartlett, Stéphane Boucheron, and Gábor Lugosi. Model selection and error estimation. *Machine Learning*, 48:85–113, 2002.

[Bur89] Prabir Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989.

[Efr79] Bradley Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.

[Efr83] Bradley Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.*, 78(382):316–331, 1983.

[Gei75] Seymour Geisser. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70:320–328, 1975.

[Kol01] Vladimir Koltchinskii. Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory*, 47(5):1902–1914, 2001.

[Kol06] Vladimir Koltchinskii. 2004 IMS Medallion Lecture: Local Rademacher Complexities and Oracle Inequalities in Risk Minimization. *Ann. Statist.*, 34(6), 2006.

[LW04] Gábor Lugosi and Marten Wegkamp. Complexity regularization via localized random penalties. *Ann. Statist.*, 32(4):1679–1697, 2004.

[Shi97] Ritei Shibata. Bootstrap estimate of Kullback-Leibler information for model selection. *Statist. Sinica*, 7(2):375–394, 1997.

Reporter: Ursula Gather