

## REGRESSIONE, FILTRI, E APPRENDIMENTO

*Sergio Decherchi, Mauro Parodi, Sandro Ridella,*

Dipartimento di Ingegneria Biofisica ed Elettronica  
Università di Genova  
Via All'Opera Pia 11/A Genova

Alla base dei metodi di apprendimento automatico, ovvero tecniche di inferenza da dati sperimentali, spesso si trovano problemi di regolarizzazione. La teoria della regolarizzazione conduce alla formulazione di modelli che derivano sia dai dati sia da una eventuale conoscenza *a priori* sull'oggetto da modellare. In termini matematici questa conoscenza è associata al termine di regolarizzazione. Il procedimento di regolarizzazione, d'altro canto, se visto come finalizzato a individuare un segnale continuo immerso nel rumore, è classico nel campo dell'elaborazione dei segnali e corrisponde, in ambito circuitale, alla progettazione di un filtro di caratteristiche adeguate.

In entrambi i casi, il problema centrale sta nella determinazione del valore da assegnare al parametro di regolarizzazione, per mezzo del quale si bilancia l'aspetto di sintesi (ovvero l'aderenza ai dati) con l'azione di regolarizzazione. La teoria presente in letteratura non offre un criterio generale per la selezione di tale parametro.

La ricerca di tale criterio, in un primo tempo analizzata tramite la media regolarizzata, è stata portata avanti analizzando la regressione.

L'analisi fin qui compiuta ha dimostrato interessanti proprietà della regressione: si è mostrato come una formulazione oracolare di regressione è possibile utilizzando una versione generalizzata della regolarizzazione di Tikhonov. Inoltre si è mostrato che tale formulazione è una generalizzazione della regolarizzazione della media.

Si è mostrato come, alla luce della regolarizzazione oracolare, il paradosso di Stein perda significato. Il risultato finale è che la miglior stima di un vettore di parametri ha le componenti indipendenti fra loro e non dipendenti come nel lavoro storico di Stein.

Alle luce di queste analisi si è proceduto a suddividere la regolarizzazione in due famiglie: la regolarizzazione debole (shrinking) e la regolarizzazione forte (Tikhonov).

Il primo approccio si dimostra largamente insufficiente per garantire una adeguata capacità di generalizzazione in problemi di apprendimento. Di fatto, le macchine che funzionano in spazi di Hilbert a nucleo riprodotto (RKHS) possono essere considerate istanze non lineari di regolarizzazione forte.

Infine, è in fase di analisi la relazione fra la regolarizzazione di Tikhonov oracolare e il filtro di 'Wiener oracolare', nella prospettiva di definire il miglior filtro lineare possibile..

### BIBLIOGRAFIA

- [1] Vapnik V, *Statistical Learning Theory*, 1998, Wiley-Interscience Pub.
- [2] Theodoros Evgeniou, Massimiliano Pontil, Tomaso Poggio, "Regularization Networks and Support Vector Machines", *Advances in Computational Mathematics*, 2000.
- [3] James W, Stein C "Estimation with quadratic loss", *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1, pp. 361–379, 1961
- [4] Rakotomamonjy A., F.R. Bach, S.Canu, Y. GrandValet, "SimpleMKL", *Journal of Machine Learning Research*, 2008 pp. 2491-2521

- [5] J.E. Moody, "The effective number of parameters: An analysis of generalisation and regularisation in nonlinear learning systems". In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Neural Information Processing Systems 4*, pp 847-854. Morgan Kaufmann, San Mateo CA, 1992.
- [6] Geisser, Seymour, 1993,. *Predictive Inference*. New York: Chapman and Hall.
- [7] MacKay D.J.C. "A practical Bayesian framework for backpropagation networks" *Neural Computation* 4(3) pp.448-472
- [8] Ridella S, Rovetta S., Zunino R. "Circular backpropagation networks for classification", *IEEE transactions on neural networks*, 1997, (8) pp. 84-97
- [9] Thompson J.R. "Some Shrinkage Techniques for Estimating the Mean", *Journal of the American Statistical Association* 63(321), 1968, pp.113-122
- [10] Borgwardt, K., A. Gretton, M. Rasch, H.-P. Kriegel, B. Schölkopf and A. Smola: "Integrating Structured Biological data by Kernel Maximum Mean Discrepancy". *Bioinformatics* 22(4), e49-e57 08 2006.
- [11] P. Liang, F. Bach, G. Bouchard, M. I. Jordan. "Asymptotically optimal regularization in smooth parametric models". *Advances in Neural Information Processing Systems (NIPS)*, 2009.