REGRESSIONE LINEARE REGOLARIZZATA: UNA REVISIONE CRITICA

Mauro Parodi, Sandro Ridella

Dipartimento di Ingegneria Biofisica ed Elettronica
Università di Genova
Via Opera Pia 11A, 16145 Genova

La regressione lineare di dati sperimentali è un procedimento assai importante sia per l'applicazione diretta dei risultati ottenuti sia quale strumento teorico che pone le basi per lo studio della regressione non lineare.

Nella sua forma generale, il funzionale da minimizzare per il calcolo dei parametri di regressione lineare contiene un termine di regolarizzazione che ha lo scopo di limitare l'influenza dell'incertezza dei dati.

Le proprietà della soluzione di questo procedimento di minimizzazione dell'errore quadratico sono state ampiamente studiate. I risultati, tuttavia, sono disseminati su un ampio insieme di lavori nei quali raramente si mira a un inquadramento e a una presentazione sistematica. La dimostrazione di alcuni risultati, addirittura, non è disponibile, nonostante il loro impiego sia assai diffuso.

Questo lavoro si sviluppa sul filo di una revisione critica dei risultati noti per la regressione lineare regolarizzata, con l'obiettivo di proporre, per ciascun risultato, una dimostrazione basata su un insieme minimo di ipotesi. In tal modo, oltre a fornire in qualche caso dimostrazioni non reperibili nella letteratura, si pone la base per considerare ciascun risultato, dal punto di vista metodologico, come tessera di un unico mosaico. La revisione conduce inoltre ad alcuni risultati di completamento importanti per le applicazioni.

Il lavoro gravita attorno a due temi principali.

Il primo è il calcolo della varianza del rumore in uscita nel caso regolarizzato. Tale calcolo poggia sul cosiddetto "numero effettivo di parametri", per il quale esistono varie definizioni in letteratura. Il trattamento sviluppato in questo lavoro considera tali definizioni e il loro significato e punta a individuare un limite superiore e uno inferiore per la varianza.

Il secondo argomento riguarda la stima corretta dell'errore di predizione per un singolo ingresso in termini dei parametri individuati per lo stimatore. Le soluzioni proposte in letteratura poggiano su approssimazioni o soppressioni di termini spesso non giustificate in modo adeguato. Anche in questo caso, il lavoro mira a proporre e motivare, per l'errore di predizione, un limite superiore e uno inferiore.

In entrambi i casi, la difficoltà principale nel proporre i risultati noti in letteratura riguarda il trattamento del termine di "bias" dovuto alla presenza del termine di regolarizzazione del funzionale: il valore esatto del bias non può essere determinato, perchè permetterebbe un intervento deterministico sulle variabili, ad esempio sull'errore di predizione. Il lavoro punta a individuare un limite superiore per il termine di bias mediante calcoli rigorosi.

Bibliografia

- 1) Draper, N.R., Smith H.: "Applied regression analysis", Wiley 1966

- 2) Moody J. E.: "The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems", NIPS 1991, pp. 847-854
- 3) Orr, M.J.L.: "Introduction to radial basis function networks", www.anc.ed.ac.uk/rbf/papers/intro.ps, 1996