

SUPPORT VECTOR MACHINES ED ANALISI DELLE SERIE STORICHE

Diego Mancuso¹

¹ Istituto di Statistica dell'Università Cattolica del Sacro Cuore

(e-mail: diego.mancuso@unicatt.it)

ABSTRACT: Lo scopo del presente lavoro è quello di segnalare le potenzialità offerte dalle *Support Vector Machines* (SVM) nell'analisi delle serie storiche. Nella prima sezione si introduce la metodologia delle SVM partendo dal problema della generalizzazione. Nella seconda sezione si mettono alla prova le capacità di generalizzazione delle SVM attraverso uno studio di simulazione e si confrontano con quelle delle reti neurali. Nella terza sezione si applicano le SVM a una serie storica ben nota in letteratura in modo da facilitare comparazioni con metodi alternativi utilizzati in altri lavori.

KEYWORDS: non-parametric regression, SVM, time series.

1 Introduzione

Il problema della regressione si caratterizza per la stima della funzione di regressione ovvero la funzione che individua il valore medio condizionato di una variabile di risposta rispetto ai valori assunti da una serie di variabili esplicative. L'approccio non parametrico a questo problema consiste nel non postulare la classe d'appartenenza della funzione di regressione, ma di compierne la stima utilizzando metodi più diretti. Le impostazioni e i paradigmi per attuare questa operazione possono essere svariati. Adottando una qualsiasi di queste impostazioni, il problema fondamentale che si presenta risulta comunque essere quello della *generalizzazione*. Con questo termine si intende la capacità di applicare efficacemente lo strumento su osservazioni diverse da quelle utilizzate per la sua stima. La capacità di generalizzazione coincide quindi con la capacità di stimare correttamente la funzione di regressione sottostante i dati campionari evitando la loro mera riproduzione (*overfitting*) e conseguendo il loro *lisciamento* ottimale.

Nel tempo, sono state proposte varie procedure per assicurare le capacità di generalizzazione a uno stimatore non parametrico. Tra queste sicuramente la più utilizzata è la *validazione incrociata*, attraverso la quale i parametri di regolarità dello stimatore sono scelti rendendo minimi dei criteri di perdita applicati a sottoinsiemi del campione temporaneamente non utilizzati per la stima degli altri parametri.

Un approccio radicalmente diverso alla questione della generalizzazione sta alla base delle cosiddette *Support Vector Machines* (SVM). A partire dal contributo di

Vapnik and Chervonenkis 1971 si è sviluppato, in ambito probabilistico, un filone di studi volto a costruire soglie PAC (*Probably Approximately Correct*). Queste soglie hanno la caratteristica di delimitare, con probabilità prefissata δ , la differenza tra la stima empirica di un funzionale e il suo valore effettivo. La maggiorazione offerta dalle soglie PAC avviene in maniera uniforme rispetto ad uno spazio H di distribuzioni di probabilità *compatibili* con le osservazioni del campione D e il valore scelto di δ . Si noti come queste soglie possono essere costruite per qualsiasi numerosità campionaria n . Ad esempio una soglia PAC, $\varepsilon(n, H, \delta)$, riferita alla stima F_n della probabilità di un certo evento A dà luogo ad un enunciato del tipo:

$$P_{\phi}^n [D : |F_n(A) - P(A)| > \varepsilon(n, H, \delta)] \leq \delta$$

La teoria delle soglie PAC si è notevolmente raffinata nei decenni successivi e una sua esposizione estesa si trova nel volume Vapnik 1998.

Dal punto di vista della regressione non parametrica, è evidente come un interesse per le soglie PAC nasce quando queste vengono riferite all'errore di generalizzazione commesso da uno stimatore. Le *Support Vector Machines* risultano quindi essere un frutto diretto dell'indagine sulle soglie PAC in quanto definibili sinteticamente come stimatori con struttura:

$$f_n(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV}^s w_i K_{\sigma}(\mathbf{x}, \mathbf{x}_i) \quad (1)$$

in cui i pesi w_i e la complessità s sono determinati minimizzando queste soglie riferite all'errore di generalizzazione. Con $K_{\sigma}(\mathbf{x}, \mathbf{z})$ si indica una funzione *Kernel* ovvero una funzione interpretabile come prodotto interno eseguito nello spazio nel quale le variabili esplicative \mathbf{x} sono trasformate attraverso una funzione $\phi(\mathbf{x})$:

$$K_{\sigma}(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) \rangle.$$

Il nucleo gaussiano $\exp(-\gamma \|\mathbf{x} - \mathbf{z}\|^2)$ è un esempio di funzione *Kernel*.

In origine le SVM sono state proposte per un contesto di classificazione fra due gruppi dove quindi la variabile risposta y equivale a una etichetta dicotomica $-1 / 1$. Uno schema di regressione, dove invece y è una variabile reale, può tuttavia essere ricondotto a uno di classificazione se viene considerato l'evento che l'errore dello stimatore $f_n(\mathbf{x})$ nello stimare la funzione di regressione $m(\mathbf{x})$, superi o meno, in valore assoluto, uno scalare θ . Per questo problema si possono applicare funzioni di perdita opportune e determinare pesi w_i e complessità s che minimizzano le soglie PAC corrispondenti a questo tipo di errore sulla popolazione.

Questa impostazione porta a risolvere un problema di ottimizzazione quadratica che, in estrema sintesi, diventa:

$$\max W(\mathbf{w}) = \sum_{j=1}^n y_j w_j - (\theta - \varepsilon) \sum_{j=1}^n |w_j| - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j (K_{\sigma}(\mathbf{x}_i, \mathbf{x}_j) + R^2 \delta_{ij})$$

$$\text{sotto il vincolo } \sum_{i=1}^n w_i = 0.$$

Per i passaggi e gli approfondimenti del caso, si rimanda ai manuali disponibili in letteratura come ad esempio Cristianini and Shawe-Taylor 2000.

2 I risultati di una simulazione

Al fine di valutare l'efficacia delle SVM nel risolvere il problema della generalizzazione e al contempo eseguire una comparazioni con strumenti alternativi, si è organizzato uno studio di simulazione.

Si è supposto di dover stimare la funzione di regressione $m(x,y) = xy/\text{sen}[1+(xy)^2]$ definita nel quadrato $-1 < x < 1$, $-1 < y < 1$, utilizzando delle osservazioni z_i legate a m dalla relazione $z_i = m(x_i, y_i) + \eta_i$, dove η_i è una componente di disturbo gaussiana a media nulla. Si è provveduto a formare tre campioni bernoulliani indicati come *training set*, *validation set* e *test set* composti rispettivamente da 400, 100 e 1000 osservazioni (x_i, y_i, z_i) . I valori campionari delle variabili esplicative (x_i, y_i) sono stati generati da distribuzioni uniformi definite nell'intervallo $[-1,1]$ in maniera indipendente. Per la deviazione standard (*D.ST*) della componente di errore η sono state considerate le tre ipotesi: *D.ST* = 0,25, *D.ST* = 0,50 e *D.ST* = 0,75.

Utilizzando questi campioni si è quindi stimata la componente di fondo m attraverso una SVM e una rete neurale MLP (NN MLP). Il modello SVM ha utilizzato come *kernel* il nucleo gaussiano. Tutti i parametri liberi dei modelli sono stati determinati minimizzando l'*averaged squared error (ASE)* sul *validation set*:

$$ASE(f_n) = n^{-1} \sum_{i=1}^n (f_n(x, y) - m(x, y))^2.$$

I risultati dell'*ASE* calcolato sul *test set* per i differenti valori della dispersione della componente di disturbo η sono riportati in tabella 1.

	<i>D.ST</i> (η)=0,25	<i>D.ST</i> (η)=0,50	<i>D.ST</i> (η)=0,75
<i>SVM</i>	0,00095	0,00752	0,01250
<i>NN MLP</i>	0,00281	0,01369	0,02089

Tabella 1. Valori dell'*ASE* sul *test set* calcolati per i due stimatori di regressione considerati e per i vari livelli della deviazione standard della componente di disturbo η .

Come si può notare, le SVM hanno offerto una maggiore accuratezza di stima rispetto al concorrente per ogni livello di *rumore* considerato. Si deve aggiungere che le soluzioni delle SVM si sono mostrate numericamente stabili e che i tempi di esecuzione dell'algoritmo e di selezione dei parametri sono stati dell'ordine di pochi minuti. Come *software* si è utilizzato il programma "libsvm" disponibile gratuitamente in rete e contenuto nella libreria "e1071" del programma R.

3 Utilizzo delle SVM nell'analisi delle serie storiche

Se si modella una serie storica, al netto di eventuali componenti stagionali, come

$$X_t = \mu(X_{t-1}, \dots, X_{t-p}) + \eta_t \quad t = 0, 1, 2, \dots,$$

un impiego immediato delle SVM per la loro analisi è dato dalla possibilità di stima non parametrica della componente di fondo μ utilizzabile a scopi di previsione.

A titolo di esempio si è eseguita questa operazione su una base dati ben nota in letteratura, quella delle “linci canadesi”, che considera il numero annuo di questi animali catturati sulle sponde del fiume MacKenzie nel periodo 1821-1934. Questa base dati è spesso utilizzata in letteratura a motivo delle relazioni autoregressive non lineari che sono presenti sulla sua trasformata logaritmica: $\{Y_t = \log_{10} X_t\}$. Questa serie è stata analizzata, tra gli altri, in Tong 1990 dove le viene applicato un modello a soglia (*Threshold*) definito su 6 ritardi temporali della serie e in Brockwell and Davis 1991 dove invece si utilizza un modello autoregressivo lineare di ordine 12 (*AR*).

Entrambi gli autori verificano le capacità previsive dei loro modelli su un *test set* costituito dalle ultime 14 realizzazioni della serie $\{Y_t\}$ ovvero quelle comprese tra gli anni 1921 e 1934. Il *test set* è escluso dalle procedure di stima dei modelli. Le prestazioni sono valutate calcolando la media quadratica dell'errore di previsione commesso:

$$s = \sqrt{\sum_{t=1}^{14} (y_{1920+t} - \hat{y}_{1920+t})^2 / 14}.$$

In tabella 2 si è aggiunto a questi risultati quello di una SVM che ha utilizzato come variabili esplicative della serie $\{Y_t\}$ le sue due immediate realizzazioni antecedenti.

	<i>AR</i>	<i>Threshold</i>	<i>SVM</i>
<i>s</i>	0,138	0,120	0,090

Tabella 2. Valori dell'errore quadratico medio di previsione *s* per i modelli considerati

Il valore ottenuto per l'indice *s* conferma l'interesse per le SVM come strumento di previsione delle serie storiche.

Bibliografia

- BROCKWELL, P., & DAVIS, R. 1990. *Time Series: Theory and Methods*. New York: Springer-Verlag.
- CRISTIANINI, N., & SHAWE-TAYLOR, J. 2000. *An Introduction to Support Vector Machines*. Cambridge: Cambridge University Press.
- GYORFI, L., KOHLER, M., KRZYSAK, A., & WALK, H. 2002. *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer-Verlag.
- TONG, H. 1990. *Nonlinear Time Series*. Oxford: Oxford University Press.
- VAPNIK, V. 1998. *Statistical Learning Theory*. New York: Wiley.
- VAPNIK, V., & CHERVONENKIS, A. 1971. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, **16**, 264-280.