BOLLETTINO
UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

THOMAS SERAFINI

Metodi del Gradiente Proiettato per problemi quadratici e loro applicazioni nell'addestramento di Support Vector Machines

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **9-A**—*La Matematica nella Società e nella Cultura* (2006), n.2 (Fascicolo dedicato alle tesi di dottorato), p. 287–290.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2006_8_9A_2_287_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2006.

Bollettino U. M. I.
La Matematica nella Società e nella Cultura
Serie VIII, Vol. IX-A, Agosto 2006, 287-290

Metodi del Gradiente Proiettato per problemi quadratici e loro applicazioni nell'addestramento di Support Vector Machines

THOMAS SERAFINI

1. – Introduzione.

La presente tesi è composta da due parti principali. La prima parte riguarda i metodi del gradiente proiettato (GPM, Gradient Projection Methods) per la soluzione di problemi di programmazione quadratica vincolata. Nonostante la loro semplicità, questi metodi hanno ricevuto recentemente molte attenzioni grazie a due miglioramenti che ne hanno aumentato notevolmente l'efficienza: l'introduzione di un passo di linesearch, che ha permesso di globalizzare la convergenza del metodo, e una nuova categoria di regole per la selezione del passo basata sulle regole di Barzilai-Borwein. L'obbiettivo della prima parte della tesi è quello di analizzare i recenti miglioramenti sia dal punto di vista delle proprietà teoriche che riguardo alle prestazioni computazionali. Le dimostrazioni di convergenza globale saranno fornite per tre diverse strategie di linesearch e, nel caso della linesearch monotona, sarà fornita una stima della velocità di convergenza. Sarà anche presentata una estensiva analisi computazionale che metterà a confronto le diverse combinazioni di regole per la selezione del passo e di linesearch. Particolare attenzione è dedicata alla descrizione di nuove scelte adattive di lunghezza del passo, che si sono mostrate particolarmente efficaci se combinate con la linesearch monotona. La prima parte della tesi si conclude con un'analisi delle proprietà di identificazione dell'insieme attivo dei metodi GPM.

La seconda parte della tesi riguarda gli aspetti numerici legati all'addestramento delle Support Vector Machines (SVM), una particolare metodologia di apprendimento automatico. L'addestramento di una SVM richiede la soluzione di un problema quadratico vincolato di grandi dimensioni e denso. A causa delle sue dimensioni, i metodi di risoluzione standard risultano scarsamente efficienti e rendono impossibile la risoluzione di problemi massivi sugli attuali calcolatori. In questo lavoro sarà presentato un algoritmo di risoluzione, chiamato *tecnica di decomposizione*, il quale suddivide il problema originale in una sequenza di sottoproblemi più piccoli. Le implementazioni esistenti della tecnica di decomposizione suddividono il problema in sottoproblemi molto piccoli, nell'ordine di $O(10)$, in modo che possano essere risolti efficacemente anche con risolutori quadratici generali. Verrà presentata un'implementazione della tecnica di decomposizione basata su problemi di dimensione nell'ordine di $O(10^3)$; per poter risolvere efficacemente questi sottoproblemi, verranno utilizzati i metodi del gradiente proiettato studiati nella prima parte della tesi. Questo approccio è chiamato Gradient Projection Decomposition Technique (GPDT) ed è confrontato con i principali software di addestramento di SVM attualmente allo stato dell'arte. La caratteristica principale del GPDT è che risulta essere particolarmente adatto ad un'implementazione parallela poichè, confrontato con gli altri software, è in grado di operare su blocchi computazionali più grossi. Saranno presentate l'implementazione parallela (chiamata PGPDT) e una sperimentazione numerica esaustiva, e sarà mostrato che con questo approccio è possibile addestrare Support Vector Machines con data set massivi in tempi ragionevoli.

2. – Metodi del Gradiente Proiettato.

In questa sezione consideriamo il problema di programmazione nonlineare

$$\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}) \tag{1}$$

dove

- $f : R^n \longrightarrow R$ è definita su un insieme aperto contenente Ω e ha derivate parziali continue;
- $\Omega \subset R^n$ è un insieme ammissibile non vuoto, chiuso e convesso definito da vincoli semplici.

Chiamiamo $\boldsymbol{g}_k := \nabla f(\boldsymbol{x})$ il gradiente della funzione obiettivo. Lo schema generico del metodo del gradiente proiettato è descritto nell'Algoritmo 1.

Diverse versioni del metodo del gradiente proiettato possono essere implementate combinando diverse regole di selezione del passo e di linesearch. Tra le regole di linesearch possiamo citare le seguenti: la linesearch monotona è definita come:

$$\lambda_k = \arg\min_{\lambda\in[0,1]} f(\boldsymbol{x}_k + \lambda \boldsymbol{d}_k)$$

e impone la decrescita della funzione obbiettivo ad ogni interazione. La linesearch nonmonotona è descritta dal seguente test di accettabilità:

$$f(\boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k) \leq f_r + \gamma \lambda_k \boldsymbol{g}_k^T \boldsymbol{d}_k, \qquad \gamma \in]0,1[, \tag{6}$$

Un contributo originale della tesi riguarda consiste nel seguente risultato di convergenza:

TEOREMA 1. – *Nel caso di una funzione obiettivo quadratica, consideriamo un metodo del gradiente proiettato in cui al punto* 3 *è utilizzata una linesearch monotona o nonmonotona e la lunghezza del passo usata al punto* 4 *è tale che*

$0 < a_{min} \leq a_k \leq a_{max}$. *Sia* $\{\boldsymbol{x}_k\}$ *la sequenza generata dall'algoritmo. Allora, ogni punto di accumulazione* $\overline{\boldsymbol{x}}$ *di* $\{\boldsymbol{x}_k\}$ *è un punto di stazionarietà vincolata.*

Algorithm 1 – Metodo del Gradiente Proiettato

STEP 1. *Inizializzazione.* Sia $\boldsymbol{x}_0 \in \Omega$, $0 < a_{\min} < a_{\max}$, $a_0 \in [a_{\min}, a_{\max}]$; sia $k = 0$.

STEP 2. *Proiezione.* Termina se $\boldsymbol{x}_k$ soddisfa un creiterio di arresto; altrimenti calcola la direzione di ricerca

(2) $$\boldsymbol{d}_k = P_{\Omega}(\boldsymbol{x}_k - a_k \boldsymbol{g}_k) - \boldsymbol{x}_k.$$

STEP 3. *Linesearch* Calcola

(3) $$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k, \qquad \lambda_k \in (0, 1],$$

dove λ_k è una lunghezza di passo determinata da una procedura di linesearch.

STEP 4. *Aggiornamento.* Calcola $a_{k+1} \in [a_{\min}, a_{\max}]$, e $k \leftarrow k + 1$ e vai al passo 2.

Algorithm 2 – GPDT (Gradient Projection Decomposition Technique)

STEP 1. *Inizializzazione.* Sia $\boldsymbol{a}^{(1)}$ un punto ammissibile per (7), sia n_{sp} un intero tale che $n \geq n_{sp} \geq 2$; poni $k \leftarrow 1$. Si scelgano arbitrariamente n_{sp} indici per l'insieme $\mathcal{B}$.

STEP 2. *Soluzione del sottoproblema quadratico.* Mediante un metodo del gradiente proiettato calcola la soluzione $\boldsymbol{a}_{\mathcal{B}}^{(k+1)}$ del problema

(4) $$\begin{aligned} &\min \ \frac{1}{2} \boldsymbol{a}_{\mathcal{B}}^T (G_{\mathcal{B}\mathcal{B}} + \tau I) \boldsymbol{a}_{\mathcal{B}} + \left(G_{\mathcal{B}\mathcal{N}} \boldsymbol{a}_{\mathcal{N}}^k - \boldsymbol{1} - \tau \boldsymbol{a}_{\mathcal{B}}^{(k)} \right)^T \boldsymbol{a}_{\mathcal{B}} \\ &\text{sub. to} \sum_{i \in \mathcal{B}} y_i a_i = - \sum_{i \in \mathcal{N}} y_i a_i^k, \\ &\qquad 0 \leq a_i \leq C \quad \forall i \in \mathcal{B} \end{aligned}$$

e poni $\boldsymbol{a}^{(k+1)} = (\boldsymbol{a}_{\mathcal{B}}^{(k+1)^T}, \ \boldsymbol{a}_{\mathcal{N}}^{(k)} T)^T$.

STEP 3. *Aggiornamento del gradiente* Aggiorna il gradiente

(5) $$g(\boldsymbol{a}^{(k+1)}) = g(\boldsymbol{a}^{(k)}) + \begin{bmatrix} G_{\mathcal{B}\mathcal{B}} \\ G_{\mathcal{N}\mathcal{B}} \end{bmatrix} \left(\boldsymbol{a}_{\mathcal{B}}^{(k+1)} - \boldsymbol{a}_{\mathcal{B}}^{(k)} \right)$$

e termina se $\boldsymbol{a}^{(k+1)}$ soddisfa le condizioni KKT.

STEP 4. *Aggiornamento* $\mathcal{B}$. Seleziona un nuovo insieme di indici $\mathcal{B}$.

---

3. – Addestramento di Support Vector Machines.

L'addestramento di una Support Vector Machine può essere ricondotto alla risoluzione del seguente problema di programmazione quadratica vincolata:

$$\begin{aligned} & \min \quad f(\boldsymbol{a}) = \frac{1}{2}\boldsymbol{a}^T G \boldsymbol{a} - \sum_{i=1}^{N} a_i \\ & \text{subj.to} \quad \boldsymbol{y}^T \boldsymbol{a} = 0 \\ & \qquad 0 \le a_i \le C, \qquad \forall i = 1, \ldots, N \end{aligned} \tag{7}$$

dove G è la matrice Hessiana densa. La risoluzione di questo problema è stata affrontata mediante una tecnica di decomposizione in cui un metodo del gradiente proiettato è stato utilizzato come risolutore interno, come descritto nell'Algoritmo 2. L'implementazione così ottenuta, chiamata GPDT, Ò stata confrontata con i principali software esistenti (SVM^{light}, LIBSVM) e ha presentato tempi di addestramento molto competitivi. Inoltre, un contributo della tesi è quella di presentare un'implementazione parallela dell'algoritmo che permette di sfuttare le risorse di calcolo di sistemi multiprocessore. Il software prodotto è disponibile alla pagina http://dm.unife.it/gpdt.

BIBLIOGRAFIA

[1] Y.H. Dai, R. Fletcher, *Projected Barzilai-Borwein methods for Large-Scale Box-Constrained Quadratic Programming*, Numerische Mathematik, **100** (2005), 21-47.

[2] T. Serafini, G. Zanghirati, L. Zanni, *Gradient Projection Methods for Quadratic Programs and Applications in Training Support Vector Machines*, Optimization Methods and Software, **20** (2005), 353-378.

[3] V.N. Vapnik, *Statistical Learning Theory*, John Wiley and sons, New York (1998).

[4] T. Serafini, L. Zanni, *On the Working Set Selection in Gradient Projection-based Decomposition Techniques for Support Vector Machines*, Optimization Methods and Software, **20** (2005), 583-596.

Dipartimento di Matematica, Università di Modena e Reggio E.
e-mail: serafini.thomas@unimo.it

Dottorato in Matematica (sede amministrativa: Modena) - Ciclo XVII
Direttore di Ricerca: Prof. Luca Zanni - Università di Modena e Reggio E.