Maşini cu vectori suport nuanțate – o nouă abordare în clasificarea datelor

Gheorghe RADU, Ştefan-Gheorghe PENTIUC

***Abstract*—Support vector machines (SVMs) have proven to be a powerful tool for classification and regression. However, there still remain some problems to be solved. One of them is that SVMs are very sensitive to outliers because of so-called overfitting problem. In this paper a new fuzzy support vector machine, based on a fuzzy distance, is proposed to deal with the problem.**

***Index terms*—Fuzzy distance, generalized Fuzzy n-Means, Support Vector Machines.**

I. INTRODUCERE

În această lucrare vom prezenta un model original de maşini nuanțate cu vectori suport. Spre deosebire de alte modele similare (vezi, de exemplu, [7]), abordarea de față se bazează pe distanța locală în raport cu o mulțime nuanțată, definită de Gerla şi Volpe [5].

În secțiunea II sunt prezentate sumar idei şi mecanisme ale SVM-urilor „clasice" (ce pot fi urmărite pe larg în [2], [9]) , iar în secțiunea III este prezentată o maşină nuanțată cu vectori suport şi problema de optimizare asociată cazului în care datele de instruire nu sunt liniar separabile. În continuare sunt trase câteva concluzii din analiza modelului teoretic prezentat.

II. MAŞINI CU VECTORI SUPORT (SVM)

A. *Instruirea SVM: un punct de vedere geometric*

Să considerăm cazul unei probleme de clasificare binară în care tipul de învățare este instruirea (antrenarea) supervizată. Astfel, fiind dată o mulțime de funcții:

$\{f_w \mid w \in W\},\ f_w : \Re^n \to \{-1,+1\}$

şi o mulțime de exemple de instruire:

$\{(x_i, y_i)\}_{i=1,2,\ldots,m}$, unde fiecare $x_i \in R^n$ reprezintă un vector de intrare (sau formă) şi fiecare $y_i \in \{-1,1\}$ reprezintă ieşirea corespunzătoare (sau eticheta), se cere a învăța funcția optimă f_w care minimizează eroarea, adică diferența dintre ieşirea dorită şi cea obținută de maşina de învățare pentru un vector de intrare. Să presupunem că avem o maşină a cărei sarcină este să învețe maparea $x_i \to y_i$.

Gheorghe RADU is with "Henri Coandă" Air Force Academy, #160, M. Viteazul street, 500183 Braşov, gh.radu@gmail.com. He is also a PhD student at "Ştefan cel Mare" University of Suceava

Ştefan-Gheorghe PENTIUC, PhD, is with "Ştefan cel Mare" University of Suceava, #13, Universității street, 720229 Suceava, pentiuc@eed.usv.ro

De fapt, maşina este definită de o mulțime de mapări posibile $x \to f(x,w)$, unde funcția *f* depinde de un parametru *w*. Se presupune că maşina este deterministă, adică: pentru un vector de intrare *x* fixat şi pentru un parametru *w* ales, funcția va avea mereu acelaşi rezultat. O alegere particulară a lui *w* va genera ceea ce se numeşte o maşină instruită (antrenată), cu care se trece la etapa de testare.

Fiind dată o mulțime de date de antrenare, să presupunem mai întâi că cele două submulțimi de vectori de intrare corespunzând etichetelor $y_i = +1$, respectiv $y_i = -1$, sunt liniar separabile. Vectorii de intrare „pozitivi" şi „negativi" vor putea fi separați de un hiperplan.

Ecuația hiperplanului de separare este:

$$\langle w, x \rangle - b = 0, \tag{1}$$

unde $w \in R^n$ este normala la hiperplan, $b \in R$ şi $\frac{|b|}{\|w\|}$ este distanța de la origine la hiperplan. Mai simplu, *w* este *orientarea* hiperplanului iar *b* este *localizarea* acestuia.

Ilustrarea acestei idei este dată în Figura 1.

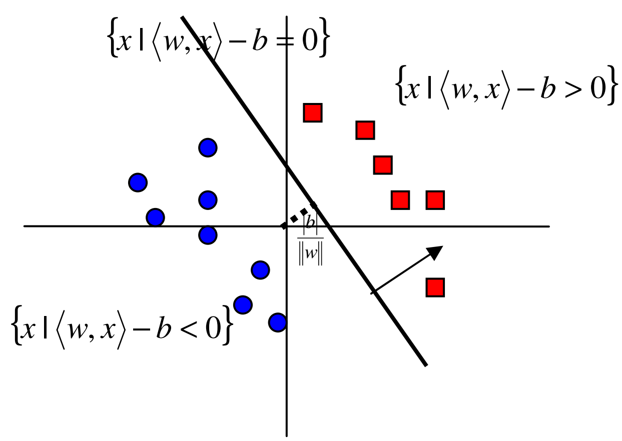

Fig. 1. Vectorii de intrare "pozitivi" şi "negativi" şi hiperplanul de separare între cele două mulțimi.

Conform [1], două submulțimi de vectori de dimensiune n sunt liniar separabile dacă şi numai dacă există $w \in R^n$ şi $b \in R$ astfel încât:

$$\begin{cases} \langle w, x_i \rangle - b \geq +1, \; y_i = +1 \\ \langle w, x_i \rangle - b \leq -1, \; y_i = -1 \end{cases}, \text{ pentru } i = 1,2,...,m \qquad (2)$$

Punctele (x_i, y_i) pentru care se realizează *egalitățile* în relațiile (1.2) se numesc **vectori suport**.

Acestea sunt punctele cele mai apropiate de suprafața de decizie. *Ştergerea acestora va modifica soluția găsită.*

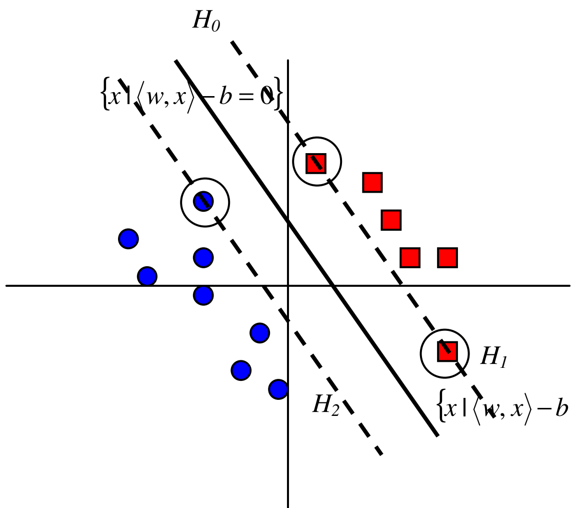

Fig. 2. Hiperplanul de separare şi hiperplanele suport pentru mulțimi separabile. Vectorii suport sunt încercuiți

Ilustrarea acestei idei este făcută în Figura 2. Hiperplanul de separare H_0 este cel care trece prin mijlocul distanței dintre cele două hiperplane paralele H_1 şi H_2 ce conțin vectorii suport pentru cele două clase.

Întrebarea este cum putem determina hiperplanul, adică w şi b , care realizează separarea optimă, adică acel hiperplan care stabileşte cea mai mare lărgime a benzii de separare (numită *margine*).

Inegalitățile (2) se pot reuni sub forma:

$$y_i(\langle w, x_i \rangle - b) - 1 \geq 0, \text{ pentru toți } i = 1,2,...,m \qquad (3)$$

Distanța de la un punct oarecare z la hiperplanul de separare w este dată de:

$$\frac{|\langle w, z\rangle - b|}{\|w\|}$$

Rezultă că distanța de la vectorii suport z_i la hiperplanul de separare este:

$$\frac{|\langle w, z_i\rangle - b|}{\|w\|} = \frac{1}{\|w\|} \qquad (4)$$

pentru toți indicii *i* pentru care $|\langle w, z_i\rangle - b| = 1$.

Aşadar, marginea este (vezi [11]) $\frac{2}{\|w\|}$ (5)

B. SVM liniare. Cazul neseparabil

Fiind dată o mulțime de date de instruire care nu este liniar separabilă, este evident că nu putem construi un hiperplan de separare fără unele erori de clasificare. Totuşi, construirea unui hiperplan optim care minimizează vectorii clasificați eronat poate fi de interes.

Discuția din paragraful precedent poate fi extinsă pentru a trata noua situație prin relaxarea restricțiilor din (2). Aceasta se poate face prin introducerea unor variabile pozitive, numite *variabile de relaxare* (*slack variables,* [3]; *slack*= *loose, relaxed*).

Introducerea acestor noi variabile „permite” unor puncte din datele de instruire o deviere de la hiperplanul lor suport, față de condiția ideală de separabilitate, de $-\frac{\xi_i}{\|w\|}$.

Restricțiile (2) devin:

$$y_i(\langle w, x_i \rangle - b) \geq 1 - \xi_i, \; i = 1,2,...,m \;, \qquad (6)$$

unde $\xi_i \geq 0$.

Astfel, pentru ca un punct din datele de instruire să reprezinte o eroare de clasificare, variabila sa de relaxare ξ_i trebuie să fie mai mare ca 1.

Punctele (x_i , y_i) pentru care în (6) avem egalități sunt vectorii suport.

În Figura 3 este reprezentată o separare liniară pentru date de instruire neseparabile.

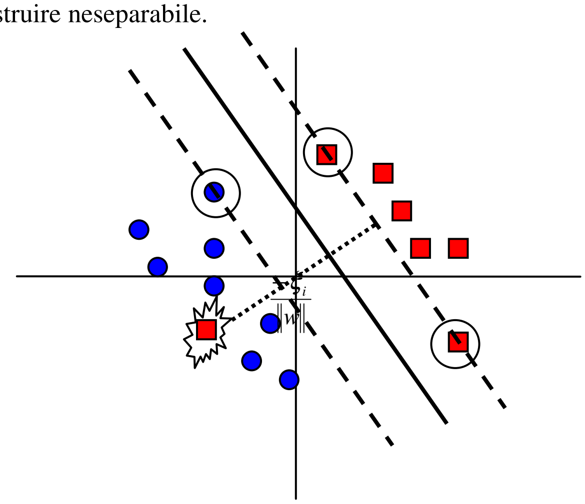

Fig. 3. Hiperplanul de separare şi hiperplanele support pentru mulțimi de date de instruire neseparabile. Vectorii suport sunt încercuiți. Este marcat şi punctul incorect clasificat.

Pentru a obține un hiperplan optimal, simultan cu (6) trebuie minimizată suma clasificărilor eronate:

$$\mathbf{min}\, C\sum_{i=1}^{m} \xi_i \,,$$

unde C este un parametru real pozitiv ales de utilizator. O valoare mai mare a lui C reprezintă o penalizare mai mare a erorilor. Astfel, problema de optimizare ia următoarea formă:

Fiind dată mulțimea de exemple de instruire

$$\{(x_i, y_i)\}_{i=1,2,...,m},$$

*să se determine valorile **optime** ale lui **w** şi **b** care **minimizează** funcția*

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i, C > 0$$

cu ***restricțiile***:

$$\begin{cases} y_i(\langle w, x_i\rangle - b) \ge 1 - \xi_i \\ \xi_i \ge 0 \end{cases}, i = 1,2,...,m.$$

Se poate vedea că această abordare satisface principiul minimizării riscului structural: dimensiunea VC (Vapnik-Cervonenkis) să fie minimizată şi datele de instruire să fie separate cu cât mai puține excepții posibil, cu satisfacerea condițiilor (5) şi (6).

Considerăm acum problema duală a SVM (a se vedea [6]) pentru acest caz; Pentru această problemă, lagrangeianul are următoarea expresie:

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\left[y_i(\langle w, x_i\rangle - b) - 1 + \xi_i\right] - \sum_{i=1}^{m}\mu_i\xi_i,$$

unde variabilele α_i şi μ_i sunt multiplicatorii lui Lagrange.

Aplicând condițiile KKTL şi explicitând termen cu termen, lagrangianul devine:

$$L(w,b,\xi,\alpha,\mu) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\xi_i - \sum_{i=1}^{m}\mu_i\xi_i$$

În consecință, obținem următoarea problemă duală:

Fiind dată mulțimea de exemple de instruire

$$\{(x_i, y_i)\}_{i=1,2,...,m},$$

*să se determine **multiplicatorii lui Lagrange** $\{\alpha_i\}_{i=1,2,..,m}$ care **maximizează** funcția obiectiv*

$$Q(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle$$

cu **restricțiile**: $\begin{cases} \sum_{i=1}^{m}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{cases}$, $i = 1,2,...,m$, $C > 0$.

Valoarea optimă w^* este calculată astfel:

$$w^* = \sum_{i=1}^{m}\alpha_i^* y_i x_i .$$

În ceea ce priveşte calculul lui *b*, acesta se poate face după cum urmează ([6]):

Dacă vom considera toate valorile $\alpha_i \ne 0$ care satisfac şi condiția $\alpha_i < C$, atunci se obține aceeaşi ecuație ca şi în cazul separabil:

$$y_i(\langle w, x_i\rangle - b) - 1 = 0$$

$$\Rightarrow b = \sum_{j=1}^{m}\alpha_j y_j\langle x_j, x_i\rangle - y_i .$$

Iarăşi, este mai bine să luăm pentru b media valorilor rezultate din toate aceste ecuații.

Acele puncte pentru care $0 < \alpha_i < C$ sunt *vectorii suport*.

III. MAŞINI CU VECTORI SUPORT NUANȚATE (FSVM)

A. *„Fuzzificarea" datelor de instruire*

Tratarea în mod egal a datelor de intruire poate conduce la fenomenul numit *overfitting* (suprapotrivire neadecvată). Apare, în mod natural, ideea de a asigna fiecărui punct un grad de apartenență corespunzător importanței sale relativ la clasa din care face parte. Din moment ce fiecărui punct x_i i s-a asignat un grad de apartenență, mulțimea datelor de instruire devine o *mulțime nuanțată.*

În felul acesta, maşinile cu vectori suport, SVM, sunt transformate în maşini *nuanțate* cu vectori suport, FSVM (Fuzzy Support Vector Machines).

Întrucât FSVM, ca şi SVM, au drept obiectiv maximizarea marginii de separare şi minimizarea erorilor de clasificare, în vederea atingerii unei înalte abilități de generalizare, propunem două abordări diferite de fuzzificare a datelor de instruire şi de formulare, în consecință, a problemei de optimizare a FSVM:

1. Margine clară (mulțime clasică), dar cu apartenența difuză a punctelor la mulțimea de instruire.
2. Margine nuanțată: regiunea de separare (marginea) M este o mulțime nuanțată peste X, notația M(x) reprezentând gradul de apartenență al lui x la regiunea de separare.

În continuare, vom prezenta prima abordare. Reamintim că obiectivele principale ale acestei abordări constau în detectarea anomaliilor şi prelucrarea datelor de instruire afectate de zgomot, luând în considerare şi informații despre calitatea mulțimii de date de antrenare. Pentru atingerea acestor obiective, pot fi avute în vedere mai multe metode şi tehnici prin care atribuim grade de apartenență/ importanță/ relevanță punctelor din mulțimea de date de instruire:

a) Utilizarea unui algoritm care să atribuie grade de apartenență pe baza importanței fiecărui punct în domeniul problemei. Pentru a păstra acea caracteristică unică (printre modelele de învățare) a SVM-urilor de a nu încorpora cunoştințe despre domeniul problemei, această metodă este puțin recomandată.

b) Utilizarea unei unelte *inteligente*, capabile să trateze diferențiat surse de date cu diferite grade de încredere. Rețeaua neuronală „Fuzzy ARTMAP cu factor de relevanță" (FAMR, a se vedea, de exemplu, [8]) ia în considerare informații despre calitatea mulțimilor de date de antrenare, poate fi folosită pentru clasificare şi este superioară modelelor concurente în prelucrarea datelor afectate de zgomot.

c) O *metodă euristică originală,* ce va fi prezentată în continuare.

Fie X={x_1, x_2, …, x_m } mulțimea vectorilor din datele de instruire. Considerăm că S este o mulțime nuanțată peste X, corespunzând la o clasă de puncte. Admitem că S conține două subclase care sunt descrise de partiția nuanțată {A_1, A_2} a lui S, unde elementele lui A_1 sunt vectorii x_i pentru care y = -1 (punctele „negative"), iar elementele mulțimii A_2 sunt punctele „pozitive"; avem:

$$A_k : X \to [0,1], \quad k = 1,2; \\ A_1(x_i) + A_2(x_i) = 1, \quad i = 1,2,...,m . \tag{7}$$

Această substructură poate fi obținută cu un algoritm de clasificare cu mulțimi nuanțate folosind, de exemplu, algoritmul *n-Medii Nuanțat Generalizat* (Generalized Fuzzy n-Means, GFNM). Pentru completitudinea prezentării, dăm mai jos acest algoritm ([4]):

Algoritmul n – Medii Nuanțat Generalizat

P1. Se inițializează o partiție nuanțată $P^1 = A_1, ..., A_n$ a mulțimii nuanțate C. Se alege eroarea admisibilă *eps.* De exemplu, *eps*=10^{-5}.

P2. Se calculează prototipul L^i al fiecărei clase A_i, folosind formula:

$$L^i = \frac{\sum_{j=1}^{p} \left(A_j\left(x^j\right)\right)^2 x^j}{\sum_{j=1}^{p} A_j^2\left(x^j\right)}, \; i = 1, ..., n; \quad j = 1, ..., p .$$

P3. Se actualizează partiția nuanțată P^1 calculând o nouă partiție P^2 ai cărei atomi se obțin cu formula:

$$A_i\left(x^j\right) = \frac{C\left(x^j\right)}{\sum_{k=1}^{n} \frac{d^2\left(x^j, L^i\right)}{d^2\left(x^j, L^k\right)}}, \; i = 1, ..., n; \quad j = 1, ..., p .$$

P4. Dacă P^2 este suficient de aproape de P^1, adică $\left\|Q^2 - Q^1\right\| < eps$,

unde Q^2 şi Q^1 sunt matricile reprezentative alte partițiilor P^2 şi P^1, iar $\|Q\|$ este norma matricei Q, atunci STOP.

În caz contrar, se pune $P^1 := P^2$ şi se merge la pasul P2.

Observații.

a) Dacă, în algoritmul de mai sus, se pune C=X, se obține algoritmul n - Medii Nuanțat.

b) Norma unei matrice Q se poate defini în mai multe feluri. Putem considera, de exemplu,

$$\|Q\| = \max_{i,j}\left|Q_{ij}\right| .$$

c) Pentru cazul nostru se iau: C=S, S=X, *n*=2*, p*=*m*, iar distanța care apare în formula de la pasul **P3** este, în conformitate cu [4,5], $d = d_C^*$.

Considerăm mulțimile (clasice) supp A_1, supp A_2, A_1^* şi A_2^*, obținute prin defuzzificarea mulțimilor nuanțate A_1, respectiv A_2:

$$A_1^* = \{x / A_1(x) > 0{,}5\}; \\ A_2^* = \{x / A_2(x) > 0{,}5\}; \tag{8}$$

În modelul de SVM propus, ***definim*** lărgimea benzii de separare nuanțate, *marginea D,* astfel:

$$D = \min_{x \in A_1^*} A_1(x) + \min_{x \in A_2^*} A_2(x) \tag{9a}$$

Prin urmare, analogul formulei (5) pentru modelul de SVM nuanțat este:

$$M = \frac{\min_{x \in A_1^*} A_1(x)}{\|w\|} + \frac{\min_{x \in A_2^*} A_2(x)}{\|w\|} = \frac{D}{\|w\|} \tag{9b}$$

Din definiția (9a) rezultă imediat:

$$\frac{1}{\|w\|} < \frac{D}{\|w\|} < \frac{2}{\|w\|} \tag{10}$$

şi, prin urmare, cerința de maximizare a marginii conduce la următoarea problemă de optimizare: $\mathbf{min}\,\frac{\|w\|^2}{D}$.

Să notăm:

$$X^* = A_1^* \cup A_2^*; \tag{11}$$

$$r = card\left(X^*\right); \tag{12}$$

$$t_i = \begin{cases} A_1(x_i), dacă\; y_i = -1 \\ A_2(x_i), dacă\; y_i = +1 \end{cases} \quad \text{i=1, 2, …, m.} \tag{13}$$

Evident, $r \le m$ şi putem considera că, printr-o renumerotare a elementelor mulțimii X, primele *r* dintre acestea vor fi elementele mulțimii X^*.

Modelul bazat pe construcția de mai sus va fi numit *Fuzzy SVM* (FSVM).

Cu aceste construcții, putem trece la formularea problemei de optimizare a FSVM-urilor.

B. Problema de optimizare a FSVM

Ca şi SVM-urile, FSVM-urile sunt construite să maximizeze marginea de separare şi să minimizeze erorile de clasificare pentru a atinge o înaltă abilitate de generalizare.

Spre deosebire de SVM-uri, FSVM-urile fuzzifică termenul de penalizare (a punctelor eronat clasificate) cu scopul de a reduce efectul datelor mai puţin relevante (anomalii sau zgomote).

În acest caz, problema de optimizare cu restricţii este definită astfel:

Fiind dată mulţimea de exemple de instruire $\{(x_i, y_i, t_i)\}_{i=1,2,...,m}$,

să se determine valorile **optime** ale lui **w** şi **b** care **minimizează** funcţia obiectiv

$$\Phi(w,\xi) = \frac{1}{D}\|w\|^2 + C\sum_{i=1}^{m} t_i\xi_i,\ C > 0$$

*cu **restricţiile**:*

$$\begin{cases} y_i(\langle w, x_i\rangle - b) \ge 1 - \xi_i \\ \xi_i \ge 0 \end{cases}, i = 1,2,...,m.$$

Pentru această problemă, lagrangeianul are următoarea expresie:

$$L(w,b,\xi,\alpha,\mu,t) = \frac{1}{D}\|w\|^2 + C\sum_{i=1}^{m} t_i\xi_i - \sum_{i=1}^{m}\alpha_i\left[y_i(\langle w, x_i\rangle - b) - 1 + \xi_i\right] - \sum_{i=1}^{m}\mu_i\xi_i$$

unde variabilele α_i şi μ_i sunt multiplicatorii lui Lagrange.

Aplicând condiţiile Karush-Kuhn-Tucker-Lagrange (KKTL), după explicitarea termen cu termen lagrangeianul devine:

$$L(w,b,\xi,\alpha,\mu,t) = \sum_{i=1}^{m}\alpha_i - \frac{D}{2}\left(1 - \frac{D}{2}\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + C\sum_{i=1}^{m} t_i\xi_i - \sum_{i=1}^{m}\alpha_i\xi_i - \sum_{i=1}^{m}\mu_i\xi_i.$$

Dezvoltând termen cu termen, se obţine:

$$\begin{aligned} L(w,b,\xi,\alpha,\mu,t) &= \sum_{i=1}^{m}\alpha_i - \frac{D}{2}\left(1 - \frac{D}{2}\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \sum_{i=1}^{m}(\alpha_i + \mu_i)\xi_i - \sum_{i=1}^{m}\alpha_i\xi_i \\ &= \sum_{i=1}^{m}\alpha_i - \frac{D}{2}\left(1 - \frac{D}{2}\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \sum_{i=1}^{m}\alpha_i\xi_i + \sum_{i=1}^{m}\mu_i\xi_i - \sum_{i=1}^{m}\alpha_i\xi_i \\ &= \sum_{i=1}^{m}\alpha_i - \frac{D}{2}\left(1 - \frac{D}{2}\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle. \end{aligned}$$

În consecinţă, obţinem următoarea problemă duală:

Fiind dată mulţimea de exemple de instruire $\{(x_i, y_i, t_i)\}_{i=1,2,...,m}$,

*să se determine **multiplicatorii lui Lagrange*** $\{\alpha_i\}_{i=1,2,..,m}$ *care **maximizează** funcţia obiectiv*

$$Q(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{D}{2}\left(1 - \frac{D}{2}\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle$$

cu **restricţiile** : $\begin{cases} \sum_{i=1}^{m}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le Ct_i \end{cases}$, $i = 1,2,...,m$, $C > 0$.

A doua restricţie este obţinută din condiţia ca $\mu_i \ge 0$, $i = 1,2,..,m$.

Valoarea optimă w^* este calculată, din nou, astfel:

$$w^* = \sum_{i=1}^{m}\alpha_i^* y_i x_i.$$

În ceea ce priveşte calculul lui ***b***, acesta se poate face după cum urmează:

Dacă vom considera toate valorile $\alpha_i \ne 0$ care satisfac şi condiţia $\alpha_i < Ct_i$, atunci, din condţiile KKTL $\Rightarrow$ $\mu_i \ne 0$, $\xi_i = 0$.

În aceste condiţii, se obţine aceeaşi ecuaţie ca şi în cazul neseparabil clasic:

$$y_i(\langle w, x_i\rangle - b) - 1 = 0$$

$$\Rightarrow b = \sum_{j=1}^{m}\alpha_j y_j\langle x_j, x_i\rangle - y_i,$$

pentru orice *i* pentru care $0 < \alpha_i < Ct_i$.

Iarăşi, este mai sigur să luăm pentru *b* media valorilor rezultate din toate aceste ecuaţii.

Acele puncte pentru care $0 < \alpha_i < Ct_i$ sunt *vectorii suport*.

Observaţii

Este clar că singura diferenţă între SVM şi FSVM apare la limitele superioare ale multiplicatorilor lui Lagrange α_i din problema duală. În cazul SVM-urilor, toţi multiplicatorii sunt mărginiţi superior de constanta *C,* în timp ce, în cazul FSVM, multiplicatorii lui Lagrange α_i au limite superioare variabile, depinzând de gradul de apartenenţă. Cu cât gradul de apartenenţă al unui punct este mai mic, cu atât mai îngustă va fi regiunea de separare determinată de hiperplanul ce trece prin α_i.

Putem înlocui pe M din (9b) cu următoarea definiţie:

$$M = \frac{\min_{x\in \sup p A_1} A_1(x)}{\|w\|} + \frac{\min_{x\in \sup p A_2} A_2(x)}{\|w\|} = \frac{D}{\|w\|}. \tag{14}$$

În acest caz, inegalitatea (10) se înlocuieşte cu :

$$\frac{0}{\|w\|} < \frac{D}{\|w\|} < \frac{2}{\|w\|}. \tag{15}$$

Înlocuirea lui m cu r, conform definiţiei (12), în problema de optimizare cu restricţii (şi în toate formulele din această secţiune) va conduce la o eliminare mult mai severă a anomaliilor şi, în consecinţă, la o reducere considerabilă a dimensiunii problemei de programare pătratică.

I. CONCLUZII

În acest articol am prezentat un nou tip de clasificator instruibil nuanțat: maşini cu vectori suport *nuanțate*, FSVM – *Fuzzy Support Vector Machines*. Acest model tratează problema *suprapotrivirii (overfitting),* care apare în mod special în clasificarea binară (cu două clase) a SVM-urilor „tradiționale" (cazul formelor liniar neseparabile). Se demonstrează reducerea suprapunerii neadecvate (*overfitting*) şi, în mod corespunzător, a erorilor de clasificare.

Avantaje ale abordarii propuse :

- se reduce numărul punctelor greşit clasificate; sunt neglijate acele puncte care sunt prea depărtate de „centrii" (prototipurile) claselor lor;
- se reduce semnificativ dimensiunea problemei de programare pătratică;
- se micşorează marginea (lărgimea benzii de separare), dar aceasta va fi definită de un număr mai mic de vectori suport, ceea ce va conduce la o ameliorare a performanțelor clasificatorului în faza de testare şi, respectiv, în faza de lucru (clasificare).

REFERINȚE

[1] R.A. Bosch, J.A. Smith, *Separating Hyperplanes and the Authorship of the Disputed Federalist Papers*, American Mathematical Monthly, Volume 105, Number 7, pp. 601-608, 1998

[2] C.J.C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, 2, pp. 121-167, 1998

[3] C. Cortes, V. Vapnik, *Support Vector Networks*, Machine Learning, 20, pp. 273-297, 1995

[4] D. Dumitrescu, *Hierarchical Pattern Classification,* Fuzzy Sets and Systems, 28, pp. 145 – 162, 1988

[5] G. Gerla, R. Volpe, *The Definition of Distance and Diameter in Fuzzy Set Theory*, Studia Univ. Babeş-Bolyai, Math., 3, pp. 21-26, 1986

[6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New Jersey, 1999

[7] H.-P. Huang, Y.-H. Liu, *Fuzzy Support Vector Machines for Pattern Recognition and Data Mining,* International Journal of Fuzzy Systems, Vol. 4, No. 3, pp. 826 – 835, September 2002

[8] L. M. Sasu, *Computational Intelligence Techniques in Data Mining*, PhD Thesis, Transilvania University of Braşov, Braşov, 2006

[9] R. Stoean, D. Dumitrescu, *Evolutionary Support Vector Machines*, Technical Report, Department of Computer Science, Faculty of Mathematics and Computer Science, "Babes-Bolyai" University, 2005

[10] V. N. Vapnik, *The Nature of Statistical Learning Theory,* Springer Verlag, New York,1995

[11] A. Zaknich, *Neural Networks for Intelligent Signal Processing*, World Scientific Publishing, Singapore, 2003