

1. Atraminių vektorių klasifikatoriaus

1.1. Vapniko-Červonenkio dimensija

Atraminių vektorių klasifikatoriai (AVK) (angl *Support Vector Machines SVM*), panaudodami mokymo duomenis (X_i, Y_i) , sukuria klasifikavimo taisykles. Pradiniai AVK variantai buvo pasiūlyti Vladimiro Vapniko 1970-ųjų pabaigoje [7]. 1990 metais jis iš Maskvos persikraustė dirbti į JAV ir kartu su savo mokiniais ir kolegomis išvystė AVK metodą. AVK grįstas Vapniko-Červonenkio dimensija [8]. Vapniko-Červonenkio dimensija skaitiškai įvertina klasifikatoriaus sudėtingumą. Tarkime tiesinis klasifikatorius gali daryti daugiau už kvadratinę klasifikacijos klaidų, tačiau jo sudėtingumas yra mažesnis ir į tai reikia atsižvelgti priimant sprendimą vertinant kuris iš dviejų klasifikatorių yra geresnis (tiesinis, paprastesnis, turintis mažiau parametrų, bet darantis daugiau klasifikavimo klaidų,

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 2 of 23

Go Back

Full Screen

Close

Quit

ar kvadratinis, sudėtingesnis, turintis daugiau parametų, bet darantis mažiau klasifikavimo klaidų). Apibrėšime formaliai Vapniko-Červonenkio dimensiją.

1 apibrėžimas. *Sakysime turime klasifikatorių f , charakterizuojamą parametų rinkiniu Θ , ir testavimo duomenis $X_1, X_2, \dots, X_N$. Sakysime, kad klasifikatorius f **atskiria** duotus duomenis X_i , jei bet kokioms binarinėms žymėms $Y_i \in \{-1, 1\}$ ¹, $i = 1, 2, \dots, N$, egzistuoja klasifikatoriaus f parametų rinkinys Θ su kuriuo visi duomenys yra teisingai klasifikuojami.*

*Klasifikatoriaus f Vapniko-Červonenkio (VC) dimensija D_{VC} vadinamas maksimalus duomenų X_i kiekis, kurį klasifikatorius **atskiria**.*

¹Aprašant atraminių vektorių klasifikatorių patogiau laikyti, kad binarinės žymės yra -1 ir 1, o ne 0 ir 1

Home Page

Title Page

◀

▶

◀

▶

Page 3 of 23

Go Back

Full Screen

Close

Quit

Pavyzdžiui, jei klasifikatorius yra tiesė plokštumoje, o klasifikuojami duomenys plokštumos taškai, tai bet kokius nekomplanarius tris plokštumos taškus su bet kokiomis binarinėmis žymėmis galime atskirti tiese, tačiau ne visada keturis plokštumos taškų galėsime atskirti tiese (žiūr. 1 iliustraciją). Vadinasi tiesinio plokštumos taškų klasifikatoriaus dimensija yra 3. Galima parodyti, kad tiesinio klasifikatoriaus klasifikuojančio D -atės erdvės taškus VC dimensija yra $D + 1$ (plokštumos taškų atveju $D = 2$ ir $D + 1 = 3$).

Vapnikas ir Červonenkis įrodė, kad jei klasifikatorius nežinomų duomenų klasifikavimo klaida gali būti įvertinta iš viršaus tokia išraiška

$$P(\text{Klasif. pagal } f) \leq P(\text{Klasif. pagal } f \text{ testavimo duom.}) + \sqrt{\left(\frac{D_{VC}(\log(2N/D_{VC}) + 1) - \log(\delta/4)}{N}\right)}. \quad (1)$$

Home Page

Title Page

◀◀

▶▶

◀

▶

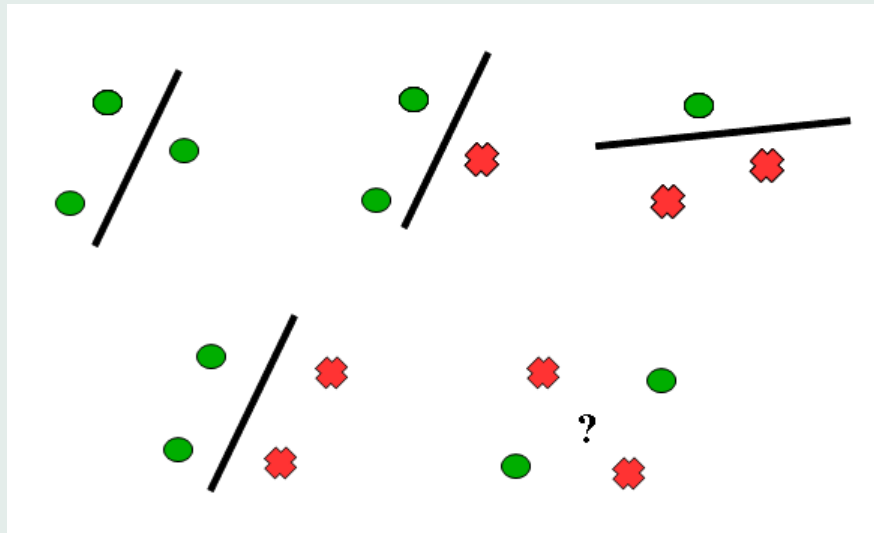
Page 4 of 23

Go Back

Full Screen

Close

Quit



1 paveikslėlis: Plokštumos taškų tiesinio klasifikatoriaus Vapniko-Červonenkio dimensija. Bet kaip sužymėtus tris taškus galime atskirti tiese, o keturis nebūtinai.

Home Page

Title Page

◀

▶

◀

▶

Page 5 of 23

Go Back

Full Screen

Close

Quit

Čia $1 - \delta$ tikimybė su kuria galioja klasifikavimo klaidos įvertis, o N - testavimo duomenų kiekis. (1) nelygybė suteikia galimybę įvertinti pasirinkto klasifikatoriaus kokybę ir nurodo galimą strategiją konstruojant klasifikatorių. Atraminių vektorių klasifikatorius yra viena tokios strategijos realizacijų.

1.2. Tiesinis atraminių vektorių klasifikatorius

Tarkime turime mokymo duomenų imtį (X_i, Y_i) , $i = 1, 2, \dots, L$, $Y_i \in \{-1, 1\}$. MNIST skaitmenų atveju $L = 60000$, o X_i gali būti $28 \times 28 = 784$ pilkumo lygmenų matrica, t. y. kiekvienas duomuo X_i yra 784 dimensijų vektorius, kurio kiekviena komponentė įgyja reikšmes pilkumo lygmenų aibėje $\{0, 1, \dots, 255\}$. Kadangi patogiau sia vizualizuoti dvimačius duomenų vektorius, tai toliau pateiksime klasifikuojamų plokštumos taškų iliustracijas. 2 paveikslėlyje pateiktas tiesė pilnai atskiriamų plokštumos taškų pavyzdys.

Home Page

Title Page

◀

▶

◀

▶

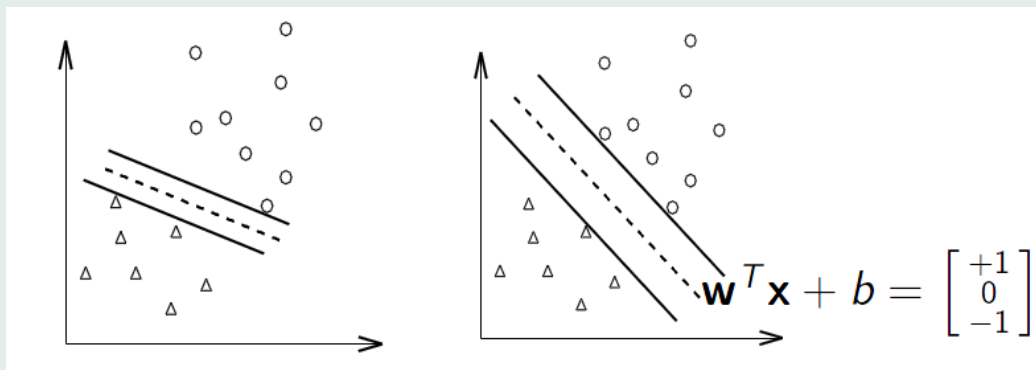
Page 6 of 23

Go Back

Full Screen

Close

Quit



2 paveikslėlis: Dviejų rūšių plokštumos taškai (skrituliukai ir trikampiai) pilnai atskiriami linija. Palyginkite du galimus iliustracijoje pateiktus tiesinės klasifikacijos variantus.

Tiesinis klasifikatorius gali būti užrašytas lygtimi

$$\mathbf{w}^T \mathbf{X} + b = 0,$$

kur T žymi vektoriaus-stulpelio transponavimo operaciją. Daugia-
mačių duomenų atveju tokia lygtis apibrėžia erdvėje hiperplokštumą

(dvimatės erdvės atveju tiesę). Klasifikuojant sieksime, kad

$$\mathbf{w}^T \mathbf{X}_i + b > 0, \text{ kai } Y_i = 1,$$

$$\mathbf{w}^T \mathbf{X}_i + b < 0, \text{ kai } Y_i = -1.$$

Todėl klasifikavimo taisyklę f testavimo duomenims $\mathbf{X}$ apibrėšime formule

$$f(\mathbf{X}) = \text{sgn}(\mathbf{w}^T \mathbf{X} + b).$$

Tarkime kokiame nors taške $\mathbf{X}_i$ išraiška $\mathbf{w}^T \mathbf{X}_i + b = 1$. Tuomet skaičius $\frac{\mathbf{w}^T \mathbf{X}_i + b}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}$ žymi hiperplokštumos (tiesės) $\mathbf{w}^T (\mathbf{X} - \mathbf{X}_i) = 0$ atstumą iki koordinatinių pradžios taško. Todėl skaičius

$$\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{(\mathbf{w}^T \mathbf{w})}}$$

lygus atstumui tarp hiperplokštumų $\mathbf{w}^T \mathbf{X} + b = \pm 1$.

Intuityviai aišku, kad tiesinis klasifikatorius, pilnai atskiriantis duomenis, tuo geresnis, kuo didesnis atstumas tarp tokių hiperplokštumų. Vapnikas ir Červonenkis pateikia tikslią atstumo tarp

taškus atskiriančių hiperplokštumų maksimizavimo interpretaciją klasifikacijos patikimumo terminais, bet mes apsiribosime tik šio principo panaudojimu. Taigi mums reikia rasti tokį svorių vektorių $\mathbf{w}$ ir poslinkį b , kad

$$\min_{\mathbf{w}, b} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w}, \quad (2)$$

$$\text{kai} \quad Y_i(\mathbf{w}^T \mathbf{X}_i + b) \geq 1, \quad i = 1, 2, \dots, L. \quad (3)$$

1.3. Tiesiškai neatskiriamų duomenų klasifikavimas atraminių vektorių metodu

Pateiktas atraminių vektorių klasifikatorius turi du akivaizdžius trūkumus.

- Jei duomenys yra bent kiek sudėtingesni, gali neatsirasti tiesinio klasifikatoriaus, kuris visiems duomenų indeksams i garantuotų (3) nelygybę. Pavyzdžiui plokštumoje galima rasti dviejų klasių

keturis taškus, kurių neįmanoma atskirti viena tiese (žiūr. 1 pvz.).

- Skirtingų žymių duomenys gali gerai atsiskirti nesudėtinga kreive, kuri nėra tiesė. 3 paveikslėlis iliustruoja tokių duomenų pavyzdį.

Kad išspręsti pirmąją problemą, 1995 metais Cortes ir Vapnikas [9] pasiūlė lanksčią (3) ir reguliarizuotą (2) lygčių versiją. Antra problema sprendžiama padidinus pradinių duomenų dimensiją, kur papildomos dimensijos sudarytos iš pradinių kintamųjų netiesinių darinių.

Netiesinį atraminių vektorių klasifikacijos uždavinį galima užrašyti taip:

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^L \xi_i, \quad (4)$$

$$\text{kai} \quad Y_i(\mathbf{w}^T \phi(\mathbf{X}_i) + b) \geq 1 - \xi_i, \quad (5)$$

$$\text{ir} \quad \xi_i \geq 0, \quad i = 1, 2, \dots, L. \quad (6)$$

Home Page

Title Page

◀◀

▶▶

◀

▶

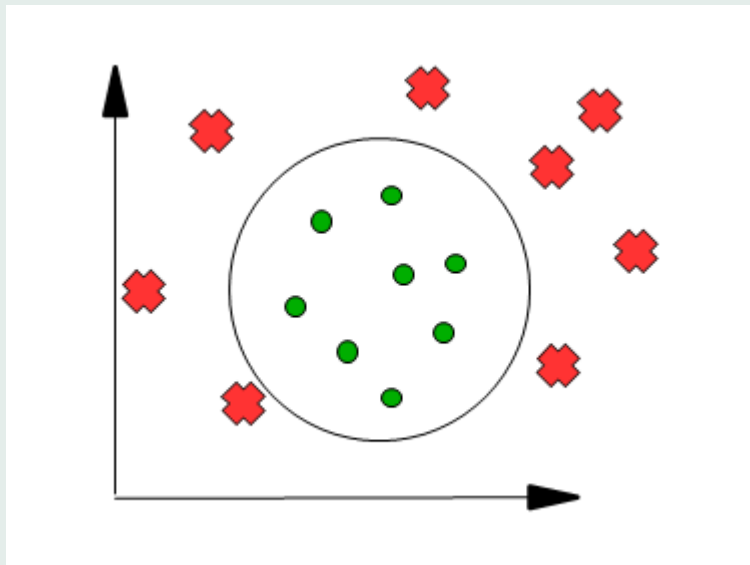
Page 10 of 23

Go Back

Full Screen

Close

Quit



3 paveikslėlis: Dviejų žymių plokštumos taškai, kuriuos lengva atskirti apskritimu, tačiau klasifikavimo bet kokia tiesė rezultatai bus labai žemos kokybės.

Home Page

Title Page

◀

▶

◀

▶

Page 11 of 23

Go Back

Full Screen

Close

Quit

Čia $C \geq 0$ yra *baudos* parametras. Jei C reikšmė didelė, kiekviena teigiama ξ_i reikšmė stipriai didina minimizuojamą funkcionalą. Todėl kuo baudos daugiklio C reikšmė didesnė, tuo labiau atsižvelgiama į blogiausiai klasifikuojamus duomenis ir atvirkščiai, kai C mažas minimizuojamo funkcionalo vertę lemia narys $\frac{1}{2}\mathbf{w}^T\mathbf{w}$. Funkcija $\phi = \phi(\mathbf{X})$ tiesinį klasifikatorių padaro netiesiniu. Pavyzdžiui, jei $X \in R^2$, tai $\phi(\mathbf{X}) \in R^6$ galima apibrėžti taip:

$$\phi(X) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2).$$

Iš vektoriaus $\phi(X)$ komponentų nesunku sudaryti bet kokio plokštumos apskritimo lygtį ir todėl nesunku rasti netiesinį atraminių vektorių klasifikatorių, kuris pilnai atskirtų 3 paveikslėlio mokymo duomenis. Bendru atveju ieškomas svorių vektorius gali būti begalinio matavimo, jei $\phi = \phi(X)$ turi be galo komponentų. Pavyzdžiui kartais vienmačiams duomenims naudojama

tokia išraiška

$$\phi(x) = \exp\left(-\frac{x^2}{2}\right)\left(1, \sqrt{\frac{1}{1!}}x, \sqrt{\frac{1}{2!}}x^2, \sqrt{\frac{1}{3!}}x^3, \dots, \sqrt{\frac{1}{n!}}x^n, \dots\right).$$

Sekančiame skyrelyje bus aiški tokio pasirinkimo motyvacija.

1.4. Atraminių vektorių klasifikatoriaus paieška sprendžiant dualų uždavinį

Minimizavimo uždaviniai dažnai sprendžiami netiesiogiai, o pirma išsprendžiami jiems dualūs uždaviniai, kurių sprendiniai panaudojami užrašant pradinio uždavinio sprendinį. Mūsų atveju (4-6) uždaviniui dualus formuluojamas taip:

$$\min_{\alpha} \quad \frac{1}{2}\alpha^T Q \alpha - \mathbf{e}^T \alpha, \quad (7)$$

$$\text{kai} \quad 0 \leq \alpha_i \leq C, i = 1, 2, \dots, L, \quad (8)$$

$$\text{ir} \quad \mathbf{Y}^T \alpha = 0. \quad (9)$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 13 of 23

Go Back

Full Screen

Close

Quit

Čia $\mathbf{e} = (1, 1, \dots, 1) \in R^L$, $Q_{i,j} = Y_i Y_j \phi(\mathbf{X}_i) \phi(\mathbf{X}_j)$. Žinant dualaus uždavinio sprendinį, atraminio vektoriaus klasifikatoriaus svoriai $\mathbf{w}$ randami taip:

$$\mathbf{w} = \sum_{i=1}^L \alpha_i Y_i \phi(X_i). \tag{10}$$

Išraiška

$$K(\mathbf{X}, \mathbf{Y}) = \phi(\mathbf{X})^T \phi(\mathbf{Y}) \tag{11}$$

vadinama *branduoliu*. Kad branduolį būtų galima efektyviai apskaičiuoti, siekiama, kad jis būtų

$$\mathbf{X}^T \mathbf{Y}$$

skaliarinės sandaugos funkcija. Pavyzdžiui, jei $X \in R^2$ ir

$$\phi(X) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2),$$

tai

$$\begin{aligned} K(\mathbf{X}, \mathbf{Y}) = \phi(\mathbf{X})^T \phi(\mathbf{Y}) &= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \\ &= (1 + x_1y_1 + x_2y_2)^2. \end{aligned}$$

Populiarūs atraminio vektoriaus klasifikatoriaus branduoliai:

$$K(\mathbf{X}, \mathbf{Y}) = \exp(-\gamma \|\mathbf{X} - \mathbf{Y}\|^2) \quad (12)$$

$$K(\mathbf{X}, \mathbf{Y}) = (1 + \mathbf{X}^T \mathbf{Y} / a)^b \quad (13)$$

Pirmasis vadinamas RBF (angl. *Radial Basis Function*), o antrasis polinominiu branduoliu. RBF branduoliui galima surasti kintamųjų pakeitimą į begalinės dimensijos erdvę, kurioje skaliarinė sandauga $\phi(\mathbf{X})^T \phi(\mathbf{Y})$ sutampa su $K(\mathbf{X}, \mathbf{Y}) = \exp(-\gamma \|\mathbf{X} - \mathbf{Y}\|^2)$. Įrodysime šį teiginį, kai $\mathbf{X} = x, \mathbf{Y} = y \in R^1$. Iš tikrųjų, kadangi pagal Teiloro

skleidinį

$$\begin{aligned}
 \exp(-\gamma \|\mathbf{X} - \mathbf{Y}\|^2) &= \exp(-\gamma(x - y)^2) = \exp(-\gamma x^2 - \gamma y^2 + 2\gamma xy) \\
 &= \exp(-\gamma x^2 - \gamma y^2) \left(\sum_{i=0}^{+\infty} \sqrt{\frac{2\gamma}{i!}} x^i \cdot \sqrt{\frac{2\gamma}{i!}} y^i \right) \\
 &= \phi(x)^T \phi(y),
 \end{aligned}$$

kur

$$\phi(x) = \exp(-\gamma x^2) \left(1, \sqrt{\frac{2\gamma}{1!}} x, \sqrt{\frac{2\gamma}{2!}} x^2, \dots, \sqrt{\frac{2\gamma}{i!}} x^i, \dots \right).$$

Polinominis branduolys atitinka baigtinio matavimo atvaizdį $\phi = \phi(\mathbf{X})$.

Branduolio savybė, kai jį galima išreikšti funkcija nuo pradinių kintamųjų skaliarinės sandaugos, vadinama "branduolio gudrybe" (angl. *kernel trick*). "Gudrybės" esmė ta, kad skaičiuojant branduolį pagal (11) išraišką galima susidurti su rimbomis skaičiavimo problemomis (pvz. begalinės $(\phi(\mathbf{X}))$ dimensijos atveju), o dėl minėtos

Home Page

Title Page

◀

▶

◀

▶

Page 16 of 23

Go Back

Full Screen

Close

Quit

branduolio savybės, jį galima skaičiuoti panaudojant pradinės dimensijos duomenis, kuri yra baigtinė ir, kaip taisyklė, žymiai mažesnė nei $\phi(\mathbf{X})$ transformacijos.

Kodėl klasifikavimo metodas vadinamas "atraminių vektorių" metodu? Taip yra todėl, kad jo klasifikavimo rezultatus lemia tik tie mokymo vektoriai, kurių $\alpha_i > 0$. Tai nesunku suprasti analizuojant pradinę tiesinę metodo versiją. Geometriškai akivaizdu, kad pilnai atskiriamų plokštumos taškų klasifikacija atraminių vektorių metodu remsis dažniausiai tik trimis mokymo taškais, kurių du bus vienos klasės, o likęs kitos klasės. Panaši tendencija išlieka ir bendru atveju, nes dauguma "lengvai" klasifikuojamų mokymo duomenų $\mathbf{X}_i$ turės $\alpha_i = 0$. Mokymo vektoriai, kurių $\alpha_i > 0$ vadinami *atraminiais*. 4 paveikslėlis iliustruoja mokymo duomenų atraminių vektorių pozicijas. Iš pateikto pavyzdžio matyti, kad nemaža mokymo duomenų dalis neįtakoja klasifikavimo rezultatų, nes, grubiai tariant, tose po-

Home Page

Title Page

◀

▶

◀

▶

Page 17 of 23

Go Back

Full Screen

Close

Quit

zicijose klasifikavimo rezultatas yra akivaizdus ir SVM klasifikatorius koncentruojasi tik į sudėtingiausius mokymo duomenų atvejus. Kuo labiau atsiskiria skirtingų žymių mokymo duomenys, tuo santykinai mažesnė dalis mokymo duomenų tampa atraminiais.

1.5. MNIST duomenų klasifikavimas SVM metodu

Kaip matyti iš pateikto aprašymo, atraminių vektorių klasifikatoriaus sukūrimas yra nelengva uždavinys. Todėl eksperimentuojant mes remsimės WEKA biblioteka, kurią sukūrė ir toliau tobulina Naujosios Zelandijos Waikato universiteto studentai ([11]). Šioje bibliotekoje yra realizuotas atraminių vektorių klasifikatorius su RBF ir polinominiais branduoliais. Deja pateikus binarizuotų 28×28 duomenų matricas WEKA-i, neužtenka kompiuterio 1.5 GB atminties, kad būtų sukurtas SVM algoritmas. Todėl tenka mažinti mokymo vektorių dimensijas. Buvo atlikti keturi eksperimentai, kuriuose naudojamas

Home Page

Title Page

◀◀

▶▶

◀

▶

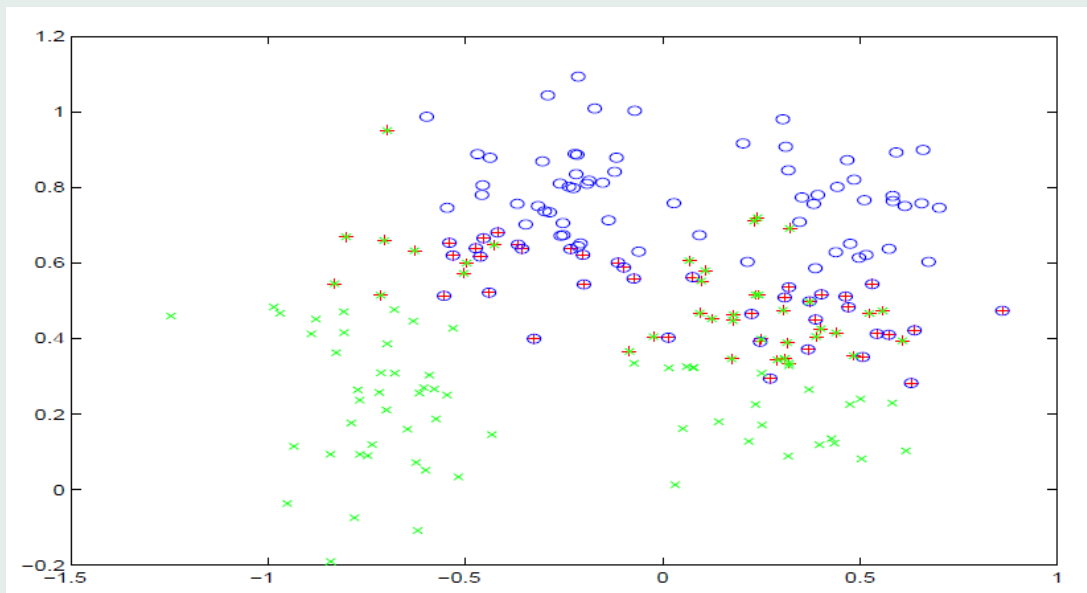
Page 18 of 23

Go Back

Full Screen

Close

Quit



4 paveikslėlis: Plokštumos taškų klasifikavimas SVM metodu. Raudonais pliusiukais pažymėti atraminiai mokymo duomenys. Iliustracija paimta iš [10]

Home Page

Title Page



Page 19 of 23

Go Back

Full Screen

Close

Quit

nutylimasis WEKA SVM klasifikatorius su skirtingomis požymių sistemomis.

- Požymiai sudarė vidurkintų skaitmenų ekstremumų kiekį telpantį į 3x3 gretimus langelius.

Gautas SVM klasifikatoriaus tikslumas: 86.3 % mokymo duomenims ir 86.2 % testavimo duomenims.

- Požymiai vidurkinto vaizdo gradiento komponentės paimtos keturis kartus išretintame tinklelyje ($28 \times 28 / 4 = 7 \times 7$).

Gautas SVM klasifikatoriaus tikslumas: 91.8 % mokymo duomenims ir 91.6 % testavimo duomenims.

- Požymiai vidurkinto vaizdo gradiento komponentės paimtos du kartus išretintame tinklelyje ($28 \times 28 / 2 = 14 \times 14$).

Gautas SVM klasifikatoriaus tikslumas: 96.38 % mokymo duomenims ir 95.43 % testavimo duomenims.

Home Page

Title Page

◀

▶

◀

▶

Page 20 of 23

Go Back

Full Screen

Close

Quit

- Požymiai gauti naudojant tikėtinumus gautus Naiviojo Bajeso metodu. Fiksuotam skaičiui randami visų skaitmenų tikėtinumai ir z -tuojo požymiu laikomas tikėtinumo, kad tai yra z skaitmuo, ir dviejų didžiausių tikėtinumų aritmetinio vidurkio skirtumas. Kiekvienam paveikslėliui tokiu būdu gauname dešimt požymių.

Gautas SVM klasifikatoriaus tikslumas: 99.84 % mokymo duomenims ir 98.16 % testavimo duomenims.

Geriausieji rezultatai gaunami paskutiniu atveju, kai požymiams naudojami Naiviojo Bajeso klasifikatoriumi gauti normuoti skaitmenų tikėtinumai. Tačiau toks rezultatas nelabai džiugina, nes klasifikuojant tiesiogiai pagal Naiviojo Bajeso tikėtinumus gaunamas 98.04 % tikslumas testavimo duomenims. Tikėtina, kad tokį nežymų pagerėjimą galima gauti paprastesniais būdais. Taip pat galima pastebėti, kad pagal mūsų eksperimentų rezultatus, gerėjant požymių

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 21 of 23

Go Back

Full Screen

Close

Quit

sistemos, gerėja klasifikacijos rezultatai, tačiau kartu stebimas ir persimokymas, pasireiškiantis ženkliai skirtumu tarp mokymo ir testavimo duomenų klasifikavimo kokybės.

Literatūra

[1] Bill Green, <http://www.pages.drexel.edu/weg22/edge.html>

[2] Kálmán Palágyi, Vengrija, <http://www.inf.u-szeged.hu/palagyi/skel/skel.html>

[3] K. Stukas, J. Janauskas, Š. Gruodis, M. Brašiškis, <http://www.mif.vu.lt/bastys/academic/ATE/skaiciai/skaiciu.atp.htm#Praktinis>

[4] D. Rutovitz, Pattern Recognition, J. Roy. Statist. Soc., vol. 129, pp. 504-530, 1966.

Home Page

Title Page

◀

▶

◀

▶

Page 22 of 23

Go Back

Full Screen

Close

Quit

[5] Feng Zhao and Xiaou Tang, CISST02 International Conference, http://mmlab.ie.cuhk.edu.hk/2002/CISST02_Fingerprint.pdf,
(Lokali kopija http://mif.vu.lt/bas-tys/academic/ATE/pirshtai/CISST02_Fingerprint.pdf)

[6] T. Y. Zhang, C. Y. Suen, A fast parallel algorithm for thinning digital patterns, Communications of the ACM, v.27 n.3, p.236-239, March 1984.

Realizacija Java kalba: <http://www.mif.vu.lt/atpazinimas/skaiciai/skelet/Zh>

[7] Vladimiro Vapniko biografija ir svarbiausi darbai, http://en.wikipedia.org/wiki/Vladimir_Vapnik

[8] Vapniko-Černovenkio dimensija, http://en.wikipedia.org/wiki/VC_dimension

[9] Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995.

Home Page

Title Page



Page 23 of 23

Go Back

Full Screen

Close

Quit

[10] Chih-Jen Lin video paskaita apie SVM,
http://videlectures.net/mlss06tw_lin_svm/, 2008.

[11] Klasifikavimo metodų biblioteka WEKA,
<http://www.cs.waikato.ac.nz/ml/weka/>