ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

Մաթեմատիկայի և մեխանիկայի ֆակուլտետ
Ֆինանսական մաթեմատիկայի ամբիոն

Կուրսային աշխատանք

Թեմա՝ Հենքային վեկտորների մեթոդ (Support Vector Machines)

Ուսանող՝ Մարիամ Սարգսյան
Ղեկավար՝ ֆ.մ.գ.թ., դոցենտ Միքայել Պողոսյան

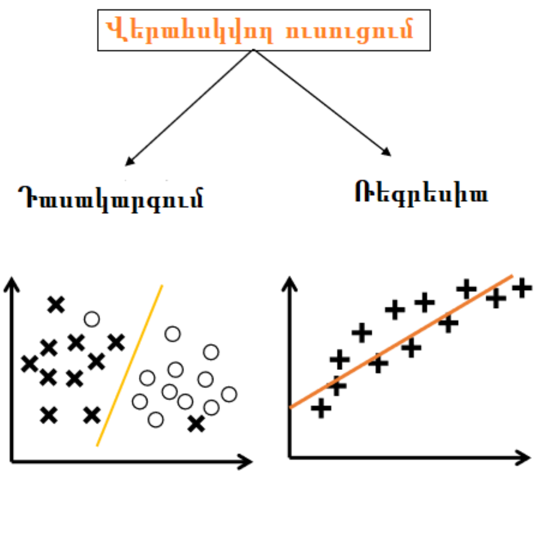
Երևան
2018

Contents

1 Ներածություն

Հենքային վեկտորների մեթոդը (Support vector machine, SVM, ՀՎՄ) տվյալների դասակարգման տեսակ է, որը համակարգչային գիտության մեջ մշակվել է 1990-ական թվականներին։ Հետազոտությունները ցույց են տալիս, որ այն մեքենայական ուսուցման արդյունավետ մեթոդ է։ Մեքենայական ուսուցումը հնարավորություն է տալիս տվյալների ուսումնասիրության հիման վրա մշակել ալգորիթմներ և դրան զուգահեռ օգտագործել վիճակագրական մեթոդներ կանխատեսումներ իրականացնելու համար։

Մեքենայական ուսուցման խնդիրները բաժանվում են 2 մեծ խմբի` վերահսկվող ուսուցման խնդիրներ (supervised Learning) և ոչ վերահսկվող ուսուցման խնդիրներ (unsupervised learning) ‘ կախված նրանից, թե տվյալների բազմությունը ներառում է արդյոք բացատրվող փոփոխականի արժեքները թե ոչ։ Վերահսկվող ուսուցման խնդիրներում տվյալների բազմությունը բաղկացած է մուտքային տվյալներից (բացատրող փոփոխական) և ելքային տվյալներից (բացատրվող փոփոխական)։ Ոչ վերահսկվող խնդիրներում տվյալների բազմությունը բաղկացած է միայն մուտքային տվյալներից և մեր կառուցող ալգորիթմն է, որ պետք է կանխատեսի մուտքային տվյալներին համապատասխանող ելքային տվյալները։



Հենքային վեկտորների մեթոդը մեքենայական ուսուցման վերահսկվող (supervised learning) ալգորիթմ է։ Այն աշխատում է հետևյալ սկզբունքով. մուքագրվում են տվյալներ, և համակարգը դրանց վերլուծության միջոցով մշակում է ալգորիթմ, որով պետք է որոշում կայացնի: Հենքային վեկտորների մեթոդը տրված 2 ուսուցողական բազմությունների համար կառուցում է նրանց բաժանող, յուրաքանչյուր բազմությունից մաքսիմալ շեմ ունեցող հիպերհարթություն: Հաճախ «հենքային վեկտորների դասակարգիչը» և «հենքային վեկտորների մեթոդը»-ը նույնացվում են, և որպեսզի խուսափենք թյուրիմացությունից առանձին կքննարկենք երկու դեպքերն էլ:

2 Ի՞նչ է հիպերհարթությունը

Երկչափ տարածության մեջ հիպերհարթությունը ուղիղ է: Եռաչափ տարածության մեջ այն հարթություն է: Երկչափ դեպքի համար այն սահմանվում է հետևյալ կերպ`

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \tag{1}$$

$\beta_0, \beta_1 \beta_2$ պարամետրերի համար: Ասելով, որ (1) հավասարումը սահմանում է հիպերհարթությունը, ի նկատի ունենք, որ եթե կամայական $X = (X_1, X_2)^T$ համար տեղի ունի (1)-ը, ապա այն հանդիսանում է կետ` ընկած այդ հիպերհարթության վրա: Հասկանալի է, որ $p > 3$ չափողականության դեպքում դժվար է այն պատկերցնել, բայց p-չափանի տարածության համար ընդհանրացված հավասարումը կարող ենք գրել հետևյալ կերպ`

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p = 0 \tag{2}$$

որը սահմանում է p-չափանի տարածությունում հիպերհարթությունը կրկին այն իմաստով, որ եթե $X = (X_1, X_2, ..., X_p)^T$ կետը p-չափանի տարածությունում բավարարում է (2)-ին, ապա X-ը ընկած է հիպերհարթության վրա: Հիմա ենթադրենք X-ը չի բավարարում (2)-ին, այսինքն X-ն ընկած է հիպերհարթության մի կողմում

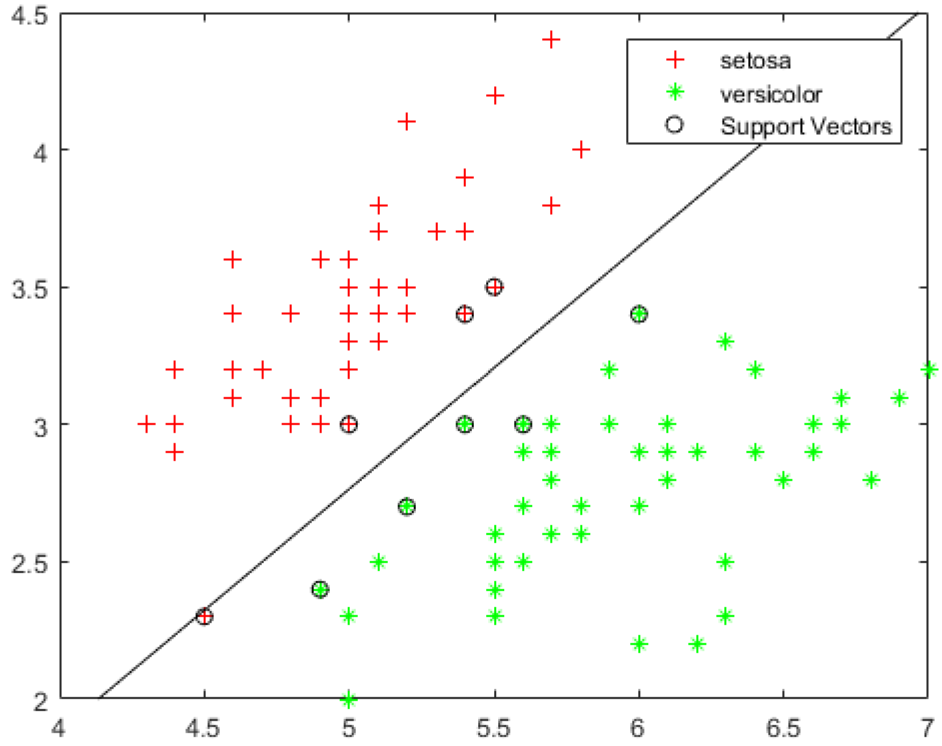
$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p > 0 \tag{3}$$

մյուս կողմից, եթե

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p < 0, \tag{4}$$

այդ դեպքում X-ն ընկած է հիպերհարթության մյուս կողմում: Այսպիսով,հիպերհարթության չափողականությունը միշտ մեկով պակաս է հիմնական տարածության չափողականությունից: Այն հիմնական տարածությունը բաժանում է երկու կիսատարածության: Երկչափ տարածությունում հիպերհարթությունը պատկերված է (նկ.1)-ում:

3 Օրինակ



(Նկ.1)

Բերված է երկչափ տարածությունում հիպերհարթության օրինակ: Այս օրինակում գրաֆիկի ստացման համար օգտագործվել է ‘Fisher’s Iris’ տվյալները: Այս դեպքում վերցրել ենք 2 տեսակի հիրիկ (setosa, versicolor), յուրաքանչյուրից 50 նմուշ, այսինքն ընդհանուր առմամբ ունենք 100 դիտարկում: Հիրիկ ծաղկի համար կատարված են չափումներ, որոնցով տեսակները կարելի է միմյանցից տարբերել: Չափված է պսակաթերթի և բաժակաթերթի երկարությունն ու լայնությունը Ենթադրենք կատարել ենք չափում և ըստ կատարված չափման ուզում ենք հասկանալ, թե ծաղիկը որ տեսակին է պատկանում: Օգտագործելով այս տվյալները՝ կառուցել ենք հիպերհարթություն, որը այս պարագայում թույլ է տալիս հասկանալ, թե ինչպես են դիտարկումները միմյանցից բաժանվում: Շրջանակների մեջ նշված են հիպերհարթությանը ամենամոտ կետերը, որոնք հետագայում կտեսնենք ինչպիսի կարևորություն ունեն:

4 Ինչպե՞ս կատարել դասակարգում՝ օգտագործելով բաժանող հիպերհարթություն

Այժմ ենթադրենք ունենք nxp չափի X մատրից, որը կազմված է p-չափանի տարածությունում n ուսուցողական դիտարկումներից,

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \quad ... \quad , x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix} \tag{5}$$

Ենթադրենք այս դիտարկումները կարող են պատկանել երկու դասի՝ 1 կամ -1, այսինքն $y_1, ..., y_n \in \{-1, 1\}$, իսկ թեստային դիտարկումներից ունենք փոփոխականների հետևյալ վեկտորը՝ $x^* = (x_1^* ... x_p^*)^T$: Մեր նպատակն է կառուցել դասակարգիչ, որը հիմնվելով ուսուցողական դիտարկումների վրա և օգտագործելով վերը նշված թեստային դիտարկումների համար կատարված չափումները, ճիշտ կդասակարգի թեստային դիտարկումները:

Ենթադրենք հնարավոր է կառուցել այնպիսի հիպերհարթություն, որը ուսուցողական դիտարկումները ամբողջովին կբաժանի իրարից: Բերենք օրինակ, որտեղ կարող ենք կառուցել այդպիսի երեք հիպերհարթություն: Նկ.2-ում կապույտ դասին պատկանող դիտարկումները որակենք որպես $y_i = 1$, իսկ վարդագույն դասին՝ $y_i = -1$: Այդ դեպքում փորձենք արտայատել բաժանող հիպերհարթությունը.

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} > 0, y_i = 1 \tag{6}$$

և

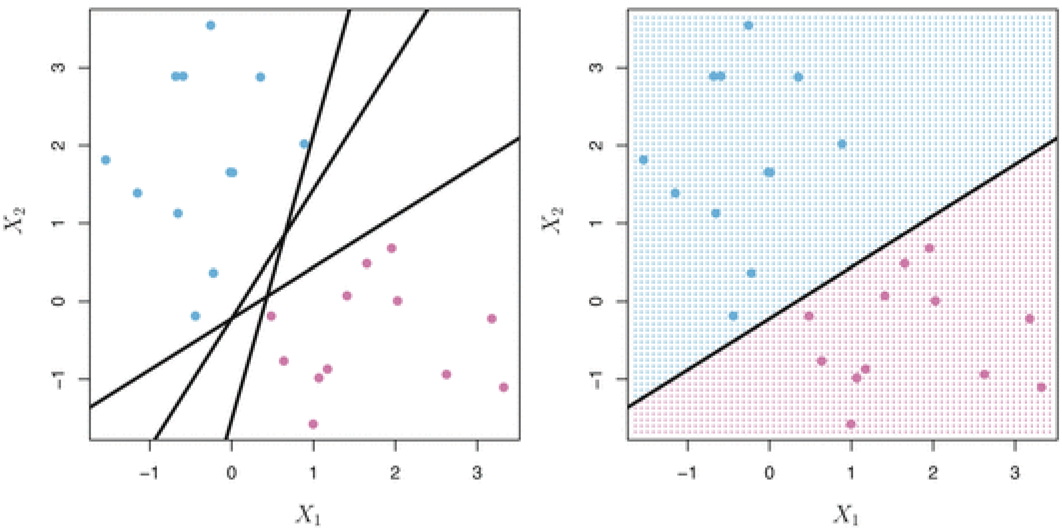
$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} < 0, y_i = -1, \tag{7}$$

որտող $\beta_0, \beta_1, ..., \beta_p$-ն բաժանող հիպերհարթության գործակիցներն են: Այս երկու պայմանները համարժեք են

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) > 0 \tag{8}$$

բոլոր i=1,...,n համար: Եթե բաժանող հիպերհարթություն գոյություն ունի, ապա այն կարող ենք օգտագործել դասակարգիչ կառուցելու համար: Եվ թեստային դիտարկումը կդասակարգենք կախված այն բանից, թե այդ հիպերհարթության որ կողմում է գտնվում, այսինքն տվյալ դեպքում դասակարգել ասելով կհասկանանք այն 1 թե -1 է: Այսինքն x^* թեստային դիտարկումը կդասակարգենք նայելով $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + ... + \beta_p x_p^*$-ի նշանին: Եթե $f(x^*)$-ը դրական է, ապա թեստային դիտարկումը 1 է և եթե $f(x^*)$-ը բացասական է, ապա այն -1 դասից է: ՈՒզում ենք, որ $f(x^*)$-ը հնարավորինս հեռու լինի 0-ից, դա կնշանակի, որ x^*-ը հիպերհարթությունից հեռու է, եթե այն մոտ լինի 0-ին, ապա այն մոտ է հիպերհարթությանը և դժվարանում է որոշելը, թե որ դասին է x^*-ը պատկանում:

4.1 Օրինակ



Հղում՝ G.James, D.Witten, T. Hastie, R.Tibshirani An Introduction to Statistical Learning

Ձախ կողմ. նշված երեք հիպերհարթությունները բաժանել են կապույտ և վարդագույն գույնով նշված դիտարկումների դասերը, որոնցից յուրաքանչյուրը չափվում է 2 փոփոխականով, այսինքն այդ փոփոխականներով որոշվում է, թե դիտարկումը որ դասին կպատկանի:

Աջ կողմ. նշված է հիպերհարությունը և կախված այդ հիպերհարթությունից կառուցվում է դասակարգիչ ըստ որի որոշվում է, թե դիտարկումը որ տեղամասին է պատկանելու՝ կապույտ թե ˜ վարդագույն:

5 Մաքսիմալ լուսանցքով դասակարգիչ (Maximal Margin Classifier)

(Նկ.2)-ի ձախ կողմից երևում է, որ նշված հիպերհարթությունները կարող են տեղափոխել փոքր-ինչ վերև կամ ներքև, աջ կամ ձախ, պտտել, այսինքն կարող են գոյություն ունենալ անվերջ թվով հիպերհարթություններ: Դասակարգիչ կառուցելու համար պետք է կարողանալ բոլոր հիպերհարթություններից հիմնավոր ընտրություն կատարել: Այդ ընտրած հիպերհարթությունը հայտնի է որպես մաքսիմալ լուսանցքով հիպերհարթություն կամ օպտիմալ բաժանող հիպերհարթություն, դա այն բաժանող հիպերհարթությունն է, որը ուսուցողական դիտարկումներից գտնվում է ամենամեծ հեռավորության վրա:

Կարող ենք հաշվարկել յուրաքանչյուր ուսուցողական դիտարկման և այդ դիտարկումները բաժանող հիպերհարթության միջև հեռավորությունը, և գտնել դրանցից մինիմալը: Այդ հիպերհարթություն-դիտարկում մինիմալ հեռավորությունը հայտնի է որպես լուսանցք:

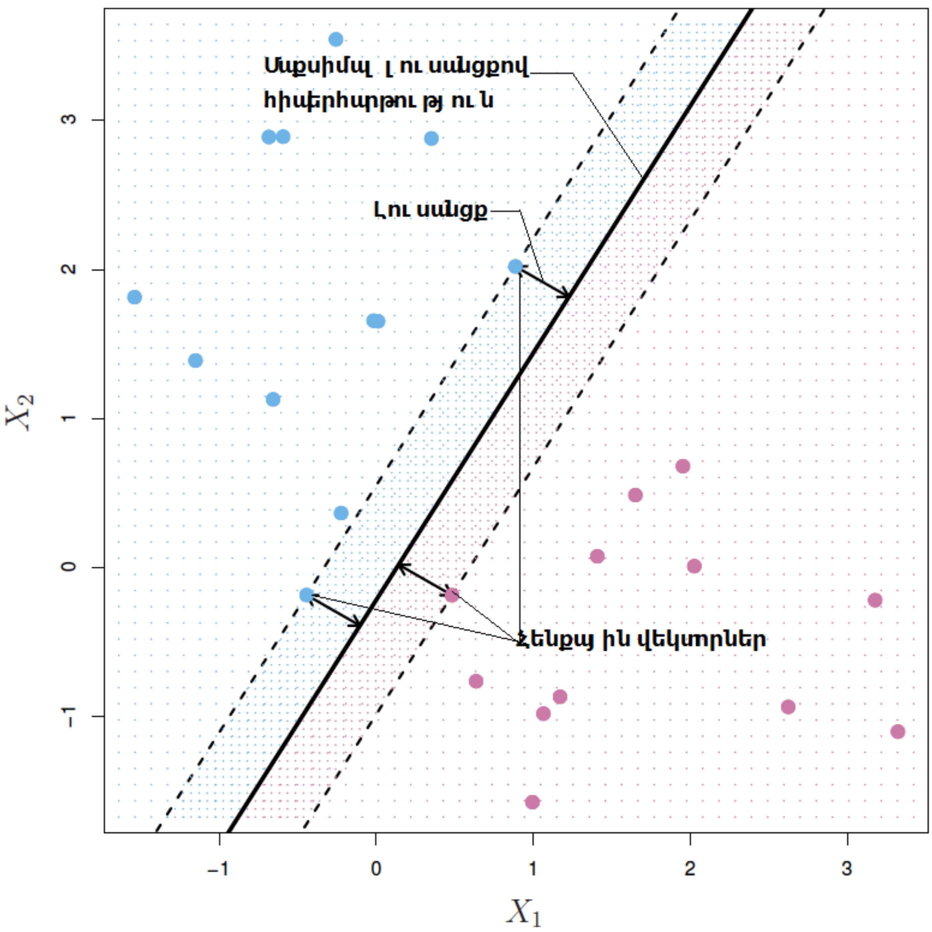
Այսպիսով՝ որպես մաքսիմալ լուսանցքով հիպերհարթություն կվերցենք բոլոր բաժանող հիպերհարթություններից ամենամեծ լուսանցք ունեցող հիպերհարթությունը: Թեստային դիտարկումը դասակարգելու համար, կօգտվենք մաքսիմալ լուսանցքով դասակարգչից, որը ըստ էության օգտագործում է այն հանգամանքը, թե այդ մաքսիմալ լուսանցքով հիպերհարթության որ կողմում է գտնվում դիտարկումը:
Մաքսիմալ լուսանցքով հիպերհարթությունը որոշեցինք ուսուցողական տվյալների համար և հույս ունենք՝ այն ճիշտ կդասակարգի նաև թեստային տվյալները:

Եթե $\beta_0, \beta_1, ..., \beta_p$-ն մաքսիմալ լուսանցքով հիպերհարթության գործակիցներն են, ապա մաքսիմալ լուսանցքով հիպերհարթությունը դասակարգում է x^* թեստային դիտարկումը՝ կախված $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + ... + \beta_p x_p^*$ արտահայտության նշանից:

Մաքսիմալ լուսանցքով հիպերհարթությունը նկարագրենք այլ կերպ՝ պատկերացնենք ունենք խոշոր «սալաքար», որը տեղադրում ենք երկու դասերի միջև, մաքսիմալ լուսանցքով հիպերհարթությունը կհանդիսանա «սալաքարի» միջին գիծ:

6 Հենքային վեկտորներ

Սահմանեցինք մաքսիմալ լուսանցքով հիպերհարթությունը. այն անմիջականորեն կախված է հենքային վեկտորներից:



(Նկ.3) Հղում՝ G.James, D.Witten, T. Hastie, R.Tibshirani An Introduction to Statistical Learning)

(Նկ.3)-ում կապույտ և վարդագույն գույներով նշված են դիտարկումների դասերը: Գրաֆիկորեն ցույց է տրված՝ ինչպես է մաքսիմում լուսանցքով հիպերհարթությունը բաժանել դրանք, և որոնք են այն դիտարկումները, որոնք հանդիսանում են հենքային վեկտորներ:

Պատկերացնենք մաքսիմալ լուսանցքով հիպերհարթության երկու կողմից լուսանցքի լայնությամբ տարված են զուգահեռ ուղիղներ, ենթադրենք կան որոշ դիտարկումներ, որոնք գտնվում են այդ ուղիղների վրա (քանի որ ուղիղները տարված են լուսանցքի լայնությամբ, դիտարկումները հավասարահեռ են հիպերհարթությունից): Այդ դիտարկումները հայտնի են որպես հենքային վեկտորներ, քանի որ p-չափանի տարածությունում նրանք վեկտորներ են: Իսկ «հենքային» բնորոշումը պայմանավորված է նրանով, որ մաքսիմալ լուսանցքով հիպերհարթության համար հանդիսանում են «հենք» այն իմաստով՝ եթե այդ կետերը որոշ չափով տեղաշարժվեն ապա մաքսիմալ լուսանցքով հիպերհարթությունը ևս կտեղաշարժվի: Իսկ մնացած դիտարկումներից որևէ մեկի տեղաշարժը չունի ազդեցություն, քանի որ հիպերհարթության լուսանցքը առաջացնում է սահմնափակում դրանց շարժի համար: Ստացվում է՝ մաքսիմալ լուսանցքով հիպերհարթությունը անմիջականորեն կախված է միայն դիտարկումների փոքր ենթաբազմությունից: Սա հենքային վեկտորների դասակարգչի և ՀՎՄ-ի կարևոր հատկություն է:

7 Մաքսիմալ լուսանցքով դասակրգիչի կառուցումը

Այժմ կառուցենք մաքսիմալ լուսանցքով հիպերհարթությունը, եթե ունենք n ուսուցողական դիտարկումներ $x_1, ..., x_n \in \mathcal{R}^p$ և այն հնարավոր դասերը, որոնց կարող են պատկանել դրանք, այսինքն $y_1, ..., y_n \in \{-1, 1\}$: Համառոտ կարող ենք ասել, որ մաքսիմալ լուսանցքով հիպերհարթությունը հետևյալ օպտիմիզացիայի խնդրի լուծումն է՝

$$maximize M,$$

$\beta_0, \beta_1, ..., \beta_p$(9)

$$\sum_{j=1}^{n} \beta_j^2 = 1 \quad (10)$$

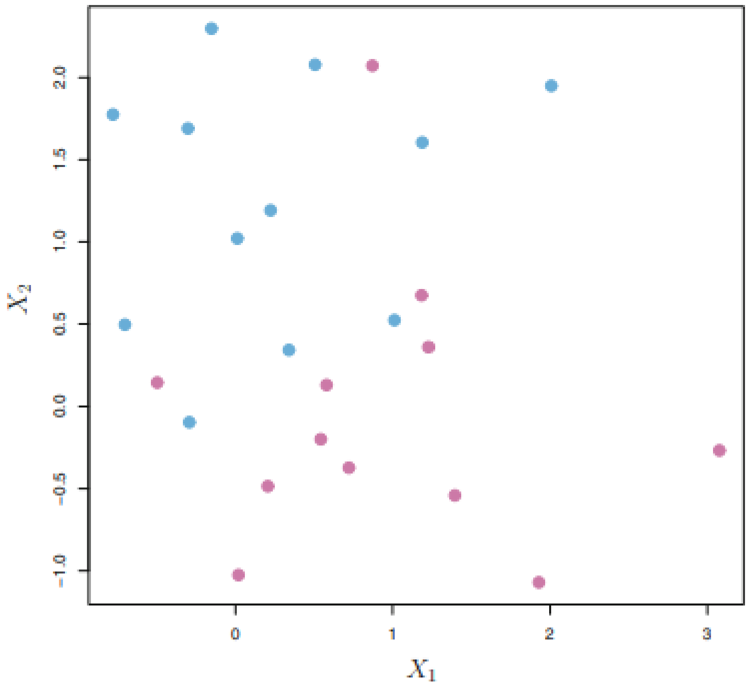
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) \geq M \ \forall i = 1, ..., n : \quad (11)$$

Նախևառաջ, բացատրենք սահմանափակումը, որը ներկայացված է հետևյալ տեսքով՝ $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) \geq M \ \forall i = 1, ..., n :$ Եթե M-ը ոչ բացասական է, ապա այդ պայմանը ապահովում է, որ յուրաքանչյուր դիտարկում լինի հիպերհարթության ճիշտ կողմում: Նշենք, որ (10)-ը իրականում սահմանափակում չէ, քանի որ հիպերհարթությունը սահմանվում է հետևյալ կերպ. $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} = 0$, որից հետևում է, որ տեղի ունի նաև հետևյալ հավասարությունը. $k(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) = 0 \ \forall k \neq 0$: Այնուամենայնիվ (10) սահմանափակումը իմաստ է հաղորդում (11)-ին և կարող ենք ցույց տալ, որ այս սահմանափակման դեպքում i-րդ դիտարկումից հիպերհարթություն հեռավորությունը տրվում է հետևյալ կերպ. $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip})$:

Այսպիսով, (10) և (11) սահմանափակումները ապահովում են, որ յուրաքանչյուր դիտարկում լինի հիպերհարթության ճիշտ կողմում և նրանից առնվազն M հեռավորության վրա: M-ը հիպերհարթության լուսանցքն է և ընտրում ենք այնպիսի $\beta_0, \beta_1, ..., \beta_p$, որպեսզի մաքսիմալացնենք այն:

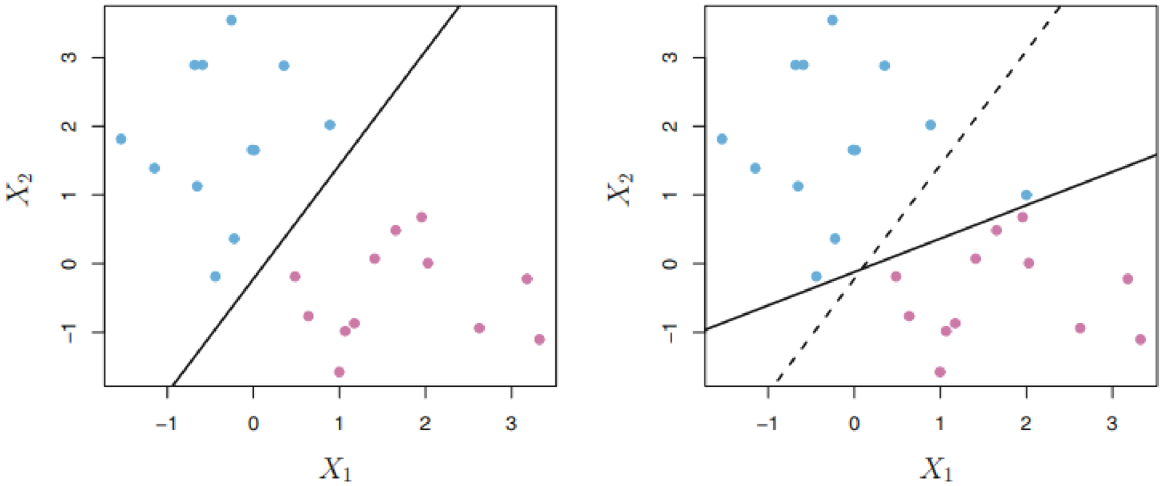
Սակայն բազմաթիվ դեպքերում, երբ դիտարկումների դասերը ամբողջապես չեն բաժանվում միմյանցից, չենք կարող օգտագործել մաքսիմալ լուսանցքով հիպերհարթությունը, այսինքն մեր խնդիրը լուծում չի ունենում: Նշված չբաժանվող դեպքը և մաքսիմալ լուսանցքով դասակարգիչի մեթոդը ընդհանրացնելով՝ մշակվել է հենքային վեկտորների դասակարգիչը:

8 Հենքային վեկտորներով դասակարգիչ (Support Vector Classifiers)



(Նկ. 4) Կապույտ և վարդագույն գույներով բերված են դիտարկումների դասեր, որոնք հիպերհարթությամբ բաժանելի չեն, այսինքն մաքսիմալ լուսանցքով դասակարգիչը կիրառելի չէ:

(Նկ. 4)-ից կարելի է տեսնել, որ դիտարկումները, որոնք պատկանում են երկու դասերին պարտադիր չէ, որ բաժանելի լինեն հիպերհարթությամբ: Փաստորեն, նույնիսկ եթե բաժանող հիպերհարթությունը գոյություն ունի, կան դեպքեր, երբ բաժանող հիպերհարթության վրա հիմնված դասակարգիչը կարող է կիրառելի չլինել: Բաժանող հիպերհարթության վրա հիմնված դասակարգիչը կատարելապես դասակարգում է բոլոր ուսուցողական դիտարկումները, այն զգայուն է յուրաքանչյուր դիտարկման նկատմամբ, այսինքն նոր դիտարկում ավելանալու դեպքում մաքսիմալ լուսանքով հիպերհարթության դիրքը կարող է կտրուկ փոփոխվել, ինչպես ցույց է տրված նկ.5-ի օրինակում, որտեղ արդյունքում տեսնում ենք, որ նոր մաքսիմալ լուսանցքով հիպերհարթությունը ունի շատ փոքր լուսանցք, ինչը խնդրահարույց է մեզ համար, քանի որ ինչպես արդեն քննարկվել է դիտարկման հեռավորությունը հիպերհարթությունից կարող է ընկալվել որպես վստահությունն այն բանի, թե դիտարկումը ինչքանով է ճիշտ դասակարգված: Ավելին, փաստը, որ առավելագույն լուսանցքով հիպերհարթությունը չափազանց զգայուն է մեկ դիտարկման փոփոխության նկատմամբ, ենթադրում է, որ այն կարող է գերհամընկնել (overfit) ուսուցողական տվյալներին:

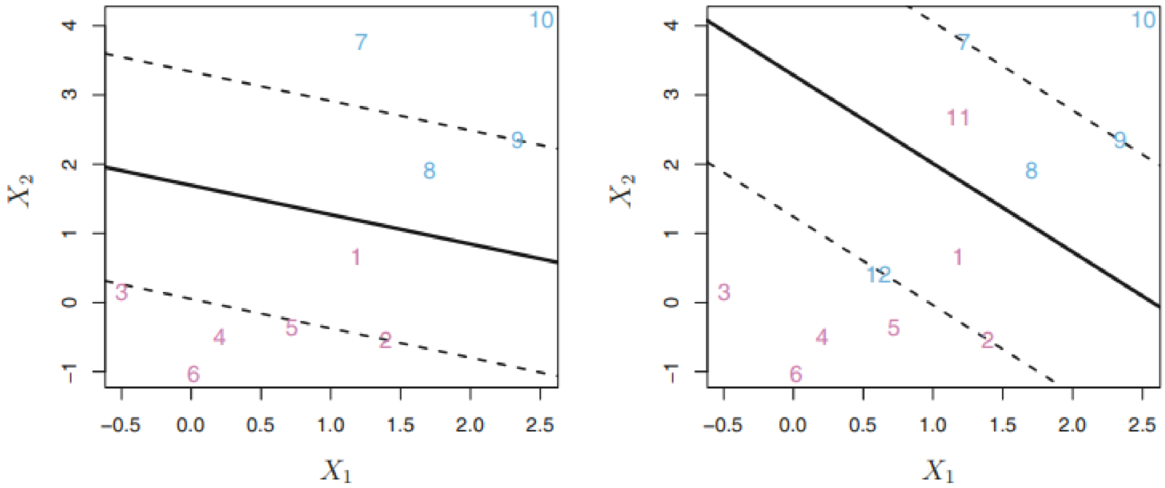


(Նկ. 5)

Այս դեպքում պետք է դիտարկել հիպերհարթության վրա հիմնված այնպիսի դասակարգիչ, որը երկու դասերը չի բաժանում կատարելապես, ի շահ՝

- յուրաքանչյուր դիտարկման անհատապես կարևորման
- ուսուցողական դիտարկումների մեծամասնության ավելի լավ դասակարգման

Այսինքն, ավելի արժեքավոր է սխալ դասակարգել մի քանի ուսուցողական դիտարկումներ՝ փոխարենը ավելի լավ աշխատանք կատարել մնացած դիտարկումները դասակարգելիս: Հենքային վեկտորների դասակարգիչը, որը հաճախ անվանում են նաև ոչ խիստ լուսանցքով դասակարգիչ (soft margin classifier) կատարում է հենց այս գործառույթը: Մաքսիմալ լուսանցք փնտրելու փոխարեն, այս մեթոդը որոշ դիտարկումների թույլ է տալիս գտնվելու լուսանցքի կամ նույնիսկ հիպերհարթության ոչ ճիշտ կողմում: (Լուսանցքը կոչվում է ոչ խիստ (soft), որովհետև այն կարող է խախտվել որոշ ուսուցողական դիտարկումների պատճառով):



(Նկ. 6)

Օրինակը ցուցադրված է Նկ. 6-ի ձախակողմյան հարթությունում: Դիտարկումների մեծ մասը լուսանցքի ճիշտ կողմում է: Այնուամենայնիվ, դիտարկումների մի փոքրիկ ենթաբազմություն գտնվում են լուսանցքի սխալ կողմում: Դիտարկումը կարող է լինել ոչ միայն լուսանցքի այլև հիպերհարթության սխալ կողմում: Փաստորեն, երբ չկա բաժանող հիպերհարթություն, այսպիսի դրությունը անխուսափելի է: Հիպերհարթության սխալ կողմում գտնվող դիտարկումները համապատասխանում են ուսուցողական դիտարկումներին, որոնց հենքային վեկտորներով դասակարգիչը սխալ է դասակարգել: Նկ. 6-ի աջակողմյան հարթությունը ներկայացնում է այսպիսի սցենար:

Հենքային վեկտորներով դասակարգիչը դասակարգում է թեստային դիտարկումը կախված նրանից, թե հիպերհարթության որ կողմում է գտնվում այն: Հիպերհարթությունը ընտրված է այնպես, որ ճշգրիտ բաժանի ուսուցողական դիտարկումների մեծ մասը երկու դասերի, այսինքն հնարավոր է նաև մի քանի դիտարկումների սխալ դասակարգում: Սա համարվում է օպտիմիզացիայի հետևյալ խնդրի լուծումը`

$$maximize M,$$

$\beta_0, \beta_1, ..., \beta_p, \epsilon_1, ..., \epsilon_n$(12)

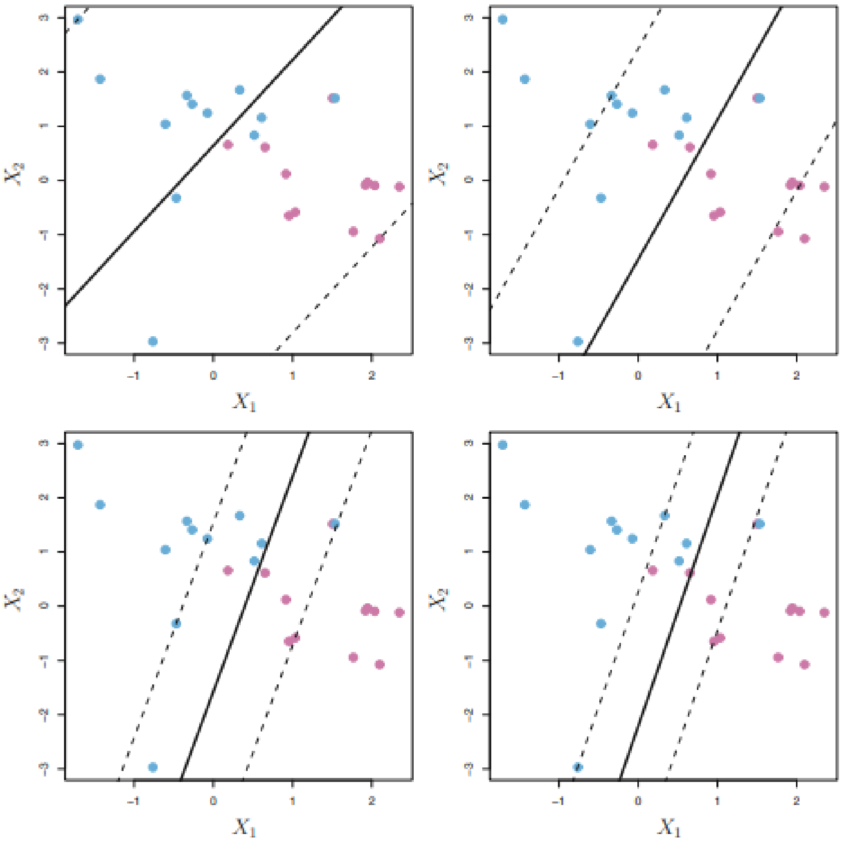
$$\sum_{j=1}^{n} \beta_j^2 = 1 \qquad (13)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \tag{14}$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C, \tag{15}$$

որտեղ C-ն ոչ բացասական կարգավորող պարամետր է (tuning parameter)։ Այստեղ նունպես M-ը լուսանցքի լայնքն է (ինչպես (11)-ում էր), և նպատակը այն հնարավորինս մեծ դարձնելն է։ (14)-ում $\epsilon_1, ..., \epsilon_n$ փոփոխականները անվանում են լրացուցիչ փոփոխականներ (slack variables), որոնք թույլ են տալիս դիտարկումներին անհատապես (individual observations) գտնվելու լուսանցքի կամ հիպերհարթության սխալ կողմում։ Երբ լուծվում է (12)-(15) խնդիրը, ինչպես նախորդ խնդրում դասակարգվում է x^* թեստային դիտարկումը, այսինքն որոշվում է, թե հիպերհարթության որ կողմում է այն գտնվում։ Այսինքն, թեստային դիտարկումը դասակարգվում է հիմնվելով՝ $f(x^*) = \beta_0 + \beta_1 x_1^* + ... + \beta_p x_p^*$-ի նշանի վրա։

Առաջին հայացքից (12)-(15) խնդիրը բարդ է թվում, սակայն կարելի է կատարել մի շարք պարզ դիտարկումներ խնդրի վարքագիծը հասկանալու նպատակով։ Առաջին հերթին, ϵ_i-ն ներկայացնում է, թե որտեղ է գտնվում i-րդ դիտարկումը՝ հիպերհարթության և լուսանցքի նկատմամբ դիրքը։ Եթե $\epsilon_i = 0$, ապա i-րդ դիտարկումը գտնվում է լուսանցքի ճիշտ կողմում, եթե $\epsilon_i > 0$, ապա i-րդ դիտարկումը գտնվում է լուսանցքի սխալ կողմում և ասվում է, որ i-րդ դիտարկումը խախտել է լուսանցքը։ Եթե $\epsilon_i > 1$, ապա այն գտնվում է հիպերհարթության սխալ կողմում։ (14)-ում C-ն սահմանափակում է ϵ_i-երի գումարը՝ որոշելով լուսանցքի (հիպերհարթության) խախտումների (violations) քանական ու ուժգնությունը։ C-ի մասին կարելի մտածել որպես հնարավորությունների ամբողջություն, որի չափով լուսանցքը կարող է խախտվել n դիտարկումների կողմից։ Եթե $C = 0$, ապա լուսանցքի խախտումների հնարավորություն գոյություն չունի, և դա $\epsilon_1 = ... = \epsilon_n = 0$ դեպքն է։ $C \neq 0$ դեպքում C-ից ոչ ավել դիտարկումներ կարող են գտնվել հիպերհարթության սխալ կողմում, որովհետև, եթե դիտարկումը գտնվում է հիպերհարթության սխալ կողմում, ապա $\epsilon_i > 1$, և պետք է տեղի ունենա $\sum_{i=1}^{n} \epsilon_i \leq C$ սահմանափակումը։ Երբ C-ն աճում է, լուսանցքի նկատմամբ խախտումները ավելի թույլատրելի են դառնում և լուսանցքը ընդլայնվում է։ Եվ ընդհակառակը, երբ C-ն նվազում է, լուսանցքի նկատմամբ խախտումները ավելի քիչ են թույլատրվում, և լուսանցքը նեղանում է։ Օրինակը ցուցադրված է (նկ.7)-ում։



(Նկ. 7)

Գործնականում C-ն կարգավորող պարամետր է, որը հիմնականում ընտրվում է cross-validation-ի միջոցով: Ինչպես այլ կարգավորող պարամետրեր՝ C-ն վերահսկում է վիճակագրական շեղում-վարիացիա (bias-variance) հակակշիռը: Երբ C-ն փոքր է, լուսանցքը նեղ է և հազվադեպ է խախտվում շնորհիվ դասակարգիչի, որը շատ լավ է համընկնում (fit) տվյալների հետ, և կարող է ունենալ ցածր շեղում (bias), բայց մեծ ցրվածք (variance): Մյուս կողմից, երբ C-ն մեծ է, լուսանցքը ավելի լայն է և ավելի շատ խախտումներ են թույլատրվում՝ տվյալների հետ դասակարգչի ավելի թույլ համընկնման (fit) պատճառով, որն այս դեպքում ավելի շեղված (biased) է, բայց կարող է ունենալ ցածր ցրվածք (variance):

(12)-(15) օպտիմիզացիայի խնդիրը ունի շատ հետաքրքիր հատկություն. պարզվում է, որ միայն այն դիտարկումները կազդեն հիպերհարթության վրա, որոնք ընկած են լուսանցքի վրա կամ խախտում են լուսանցքը, հետևաբար, հենց դրանք կազդեն կառուցված դասակարգիչի վրա: Այլ կերպ ասած, դիտարկումը, որը ընկած է լուսանցքի ճիշտ կողմում չի ազդում հենքային վեկտորներով դասակարգչի (support vector classifier) վրա: Դիտարկումները, որոնք ուղղակիորեն գտնվում են լուսանցքի վրա, կամ լուսանցքի սխալ կողմում իրենց դասի համար, հայտնի են ինչպես հենքային վեկտորներ (support vectors):

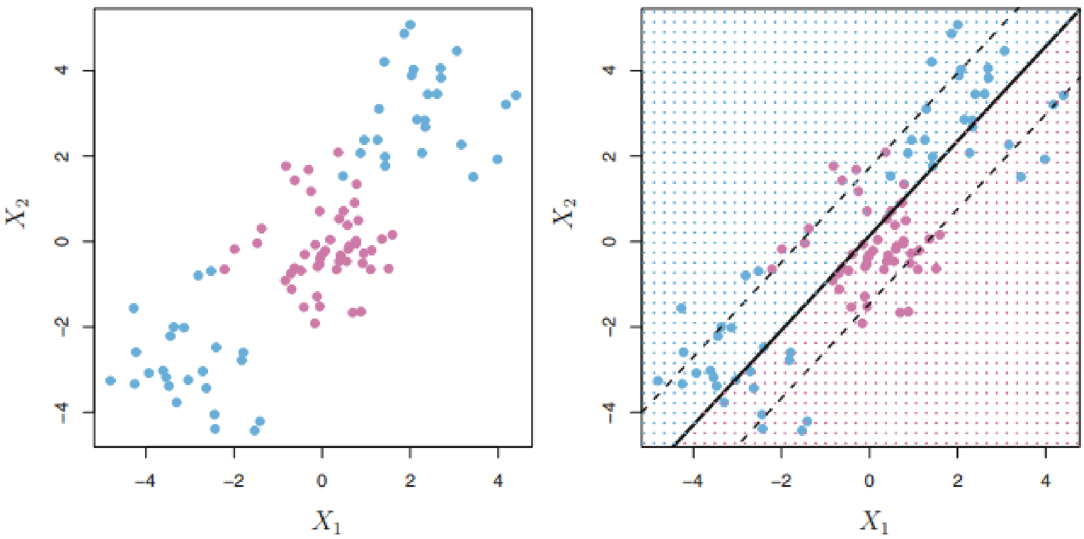
Այն փաստը, որ միայն հենքային վեկտորներն են ազդում դասակարգիչի վրա համապատասխանում է այն ենթադրությանը, որ C-ն վերահսկում է հենքային վեկտորներով դասակարգչի շեղում-ցրվածք հակակշիռը: Երբ կարգավորող պարամետր (tuning parameter) C-ն մեծ է, հետևաբար լուսանցքը ևս մեծ է, շատ դիտարկումներ խախտում են լուսանցքը, հետևաբար կան մեծ թվով հենքային վեկտորներ: Ընդհակառակը, փոքր C-ի դեպքում կան ավելի քիչ թվով հենքային վեկտորներ, արդյունքում դասակարգիչը կունենա ցածր շեղում, բայց մեծ ցրվածք

ՆԿ.7-ի վերին ձախակողմյան նկարում ներկայացված է օրինակ, որտեղ դասակարգիչը ունի փոքր ցրվածք (քանի որ շատ դիտարկումներ ներկայացնում են իրենցից հենքային վեկտորներ), բայց պոտենցիալ մեծ շեղում (C-ն մեծ է):

ՆԿ.7-ի ներքին աջակողմյան նկարում բերված է փոքր C-ի օրինակ՝ ընդամենը 8 հենքային վեկտորով:

9 Հենքային վեկտորների մեթոդ

Ինչպես տեսանք, հենքային վեկտորներով դասակարգիչը դասակարգման բնական ճանապարհ է, եթե երկու դասեր բաժանող սահմանագիծը գծային է, սակայն պրակտիկայում հաճախ ենք առնչվում նաև ոչ գծային դեպքին: Օրինակ, նկ.8-ի ձախ կողմից երևում է, որ հենքային վեկտորներով դասակարգիչը կամ այլ գծային դասակարգիչ այստեղ չի կարող ճշգրիտ աշխատել, և իսկապես աջ կողմի նկարում տեսնում ենք որ այն այս դեպքում անօգուտ է:



(Նկ. 8) Վերոնշյալ ոչ գծային դեպքի լուծման առաջարկ է փոփոխականների տարածության ընդլայնումը՝ օգտագործելով քառակուսային, խորանարդային և նույնիսկ ավելի բարձր կարգի ֆունկցիաներ: Օրինակ, փոխանակ հենքային վեկտորներով դասակարգիչը կիրառենք՝ օգտագործելով p փոփոխական.

$$X_1, X_2, ..., X_p, \tag{16}$$

այն կիրառում ենք՝ օգտագործելով 2p փոփոխական.

$$X_1, X_1^2, X_2, X_2^2, ..., X_p, X_p^2 : \tag{17}$$

Հետևապես, (12)-(15) խնդիրը այդ դեպքում կդառնա՝

$$maximize M,$$

$\beta_0, \beta_{11}, \beta_{12}, ..., \beta_{p1}, \beta_{p2}, \epsilon_1, ..., \epsilon_n$(18)

$$y_i(\beta_0 + \sum_{j=1}^{p} \beta_{j1} x_{ij} + \sum_{j=1}^{p} \beta_{j2} x_{ij}^2) \geq (M(1 - \epsilon_i)) \tag{19}$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C, \sum_{j=1}^{p} \sum_{k=1}^{2} \beta_{jk}^2 = 1 \tag{20}$$

Այս նոր ընդլայնված տարածությունում դասերն իրարից բաժանող սահմանագիծը գծային է, սակայն փոփոխականների սկզբնական տարածությունում այդ ուղիղը ունի $q(x) = 0$ տեսքը, որտեղ

q-ն 4-րդ կարգի բազմանդամ է, որի լուծումները ընդհանուր առմամբ գծային չեն: Իհարկե, կարելի է փոփոխականների տարածությունը ընդլայնել՝ օգտվելով ավելի բարձր կարգի բազմանդամից կամ $X_j X j'$ փոխազդեցությունից, որտեղ $j \neq j'$: Հասկանալի է, որ կան բազմաթիվ այլընտրանքային հնարավոր տարբերակներ ընդլայնելու փոփոխականների տարածությունը, և պետք է ուշադիր լինել, քանի որ կարող է ստացվել փոփոխականների հսկայական քանակություն, որի պատճառով հաշվարկները կդառնան անկառավարելի:

Հենքային վեկտորների մեթոդը հնարավորություն է ընձեռում ընդլայնելու փոփոխականների տարածությունը՝ օգտագործելով հենքային վեկտորներով դասակարգիչը, որով կարող ենք կատարել արդյունավետ հաշվարկներ: Ինչը նշանակում է, որ ՀՎ մեթոդը գաղափարապես ՀՎ դասակարգիչի ընդարձակումն է, որը ստացվում էր փոփոխականների տարածության ընդլայնման հետևանքով:

Փոփոխականների տարածության ընդլայնումը կկատարենք հատուկ եղանակով՝ օգտագործելով միջուկի գաղափարը (kernels): Այսպիսով, մենք ուզում ենք ընդլայնել փոփոխականների տարածությունը, որպեսզի գտնենք այն ոչ գծային սահմանը, որը բաժանում է դասերն իրարից: Միջուկի (kernel) գաղափարը պարզ և արդյունավետ եղանակ է դա անելու համար:

ՀՎ դասակարգիչը ներկայացնելիս չքննարկեցինք դրա հաշվարկը, քանի որ մանրամասները տեխնիկական բնույթ ունեն: Այնուամենայնիվ, պարզվում է, որ ՀՎ դասակարգչի (12)-(15) խնդիրը ներառում է միայն դիտարկումների ներքին արտադրյալները (inner products), փոխանակ ներառելու հենց իրենց՝ դիտարկումներին: Երկու a և b r-վեկտորների ներքին արտադրյալը սահմանվում է հետևյալ կերպ՝ $\langle a,b\rangle = \sum_{i=1}^{r} a_i b_i$:

Այսպիսով, երկու x_i և $x_{i'}$ դիտարկումների ներքին արտադրյալը (inner products) տրվում է հետևյալ կերպ՝

$$\langle x_i, x_{i'}\rangle = \sum_{j=1}^{p} x_{ij} x_{i'j} : \tag{21}$$

Եվ կարող ենք ցույց տալ, որ

- գծային ՀՎ դասակարգիչը կարող է ներկայացվել որպես՝

  $$f(x) = \beta_0 + \sum_{i=1}^{n} a_i \langle x, x_{i\rangle}, \text{(22)}$$

  որտեղ n-ը a_i պարամետրերի քանակն է, $i = 1, ..., n$ ուսուցողական դիտարկումների համար

- $a_1, ..., a_n \beta_0$ պարամետրերը գնահատելու համար մեզ պետք է $n(n-1)/2$ հատ $x_i, x_{i'}$ ներքին արտադրյալ բոլոր զույգ ուսուցողական դիտարկումների միջև

  ($n(n-1)/2$-ը զույգերի քանակն է, որը կարող ենք կազմել n դիտարկումներից)

Նշենք, որ (22)-ում, $f(x)$ ֆունկցիայի արժեքը գտնելու համար, պետք է հաշվարկել նոր կետի՝ x-ի և յուրաքանչյուր x_i ուսուցողական դիտարկման ներքին արտադրյալը: Այնուամենայնիվ, պարզվում է, որ a_i-ն զրո չէ միայն այն դեպքում, երբ ուսուցողական դիտարկումը հենքային վեկտոր չէ, հակառակ դեպքում այն հավասար է զրոյի: Այսպիսով, եթե այդ հենքային կետերի բազմությունը նշանակենք S-ով, ապա (22) տեսքի լուծման յուրաքանչյուր ֆունկցիա կարող ենք գրել որպես՝

$$f(x) = \beta_0 + \sum_{i \in S} a_i \langle x, x_{i \rangle}, \quad (23)$$

որը ավելի պարզ տեսքի է, քան (22)-ը: Այսպիսով, (22) արտահայտությունը աջ մասում պարունակում է (21)-ը, որը կփոխարինենք ներքին արտադրյալի ընդհանրացված տեսքով՝

$$K(x_i, x_{i'}), \quad (24)$$

որտեղ K-ն ֆունկցիա է, որը կանվանենք միջուկ:

Միջուկը ֆունկցիա է, որը չափում է երկու դիտարկումների նմանությունը: Օրինակ, կարող ենք գրել՝

$$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j}, \quad (25)$$

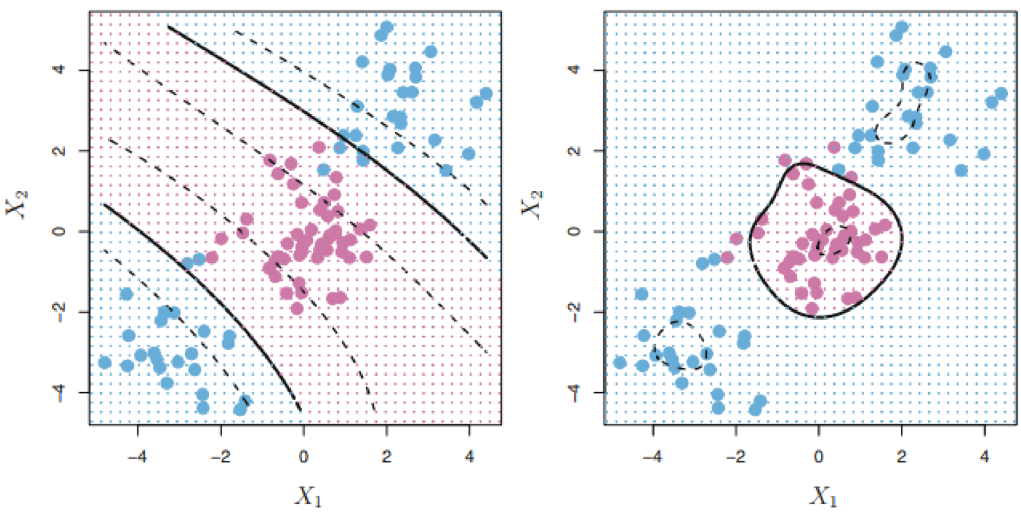
որը հենց ՀՎ դասակարգիչն է: (25)-ը հայտնի է որպես գծային միջուկ, քանի որ Հվ դասակարգիչը գծային է: Գծային միջուկը չափում է դիտարկումների զույգերի նմանությունը՝ օգտագործելով Պիրսոնի (ստանդարտ) կորելացիան: Սակայն (24)-ը կարելի է ներկայացնել նաև այլ տեսքով, օրինակ, կարող ենք գրել՝

$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d : \quad (26)$$

Այս տեսքը հայտնի է որպես d աստիճանի պոլինոմիալ միջուկ (polynomial kernel), որտեղ d-ն դրական ամբողջ թիվ է: Ստանդարտ գծային միջուկի փոխարեն օգտագործելով այսպիսի միջուկ, որտեղ $d > 1$՝ ՀՎ դասակարգիչը հանգեցնում է ավելի ճկուն որոշումների սահմանագծի (decision boundary) կառուցմանը: Երբ ՀՎ դասակարգիչը համակցվում է ոչ գծային միջուկի հետ, ինչպիսին (26)-ն է, արդյունքում ստանում ենք դասկարգիչ, որը հայտնի է որպես Հենքային վեկտորների մեթոդ: Նշենք որ այս մեթոդի դեպքում ոչ գծային Ֆունկցիան ունի հետևյալ տեսքը՝

$$f(x) = \beta_0 + \sum_{i \in S} a_i K(x, x_i) \tag{27}$$

(ՆԿ.9)-ում ձախ կողմում բերված է պոլինոմիալ միջուկով ՀՎ մեթոդի օրինակ, (նկ.8)-ի տվյալների հիման վրա, որոնք մեզ դասակարգել չէր հաջողվել:



(Նկ. 9) Ներկայացնենք ոչ գծային միջուկի մեկ այլ օրինակ՝ ռադիալ միջուկ (radial kernel), որն ընդունում է հետևյալ տեսքը՝

$$K(x_i, x_{i'}) = exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2), \tag{28}$$

որտեղ γ-ն դրական հաստատուն է: (ՆԿ.9)-ի աջ կողմում ՀՎ մեթոդը ռադիալ միջուկով ներկայացնելու օրինակ է:

Պարզենք, թե ինչպես է աշխատում ռադիալ միջուկը: Եթե տրված թեստային դիտարկումը՝ $x^* = (x_1^* ... x_p^*)^T$-ը հեռու է գտնվում x_i ուսուցողական դիտարկումից (Էվկլիդյան հեռավորություն), ապա $\sum_{j=1}^{p}(x_{j^*} - x_{ij}^*)^2$-ն մեծ է և $K(x^*, x_i) = exp(-\gamma \sum_{j=1}^{p}(x_{j^*} - x_{ij}^*)^2)$-ը շատ փոքր է, ինչը նշանակում է (27) արտահայտության մեջ x_i-ն չի ազդում $f(x^*)$-ի վրա: Հիշենք, որ x^* թեստային դիտարկման համար կանխատեսվող պիտակը (label) կախված էր $f(x^*)$-ի նշանից: Այլ կերպ ասած, այն ուսուցողական դիտարկումները, որոնք հեռու են x^*-ից դերակատարություն չունեն x^*-ի համար դասի պիտակ կանխատեսելու գործում (predicted class label): Սա նշանակում է ռադիալ միջուկը ունի շատ լոկալ վարքագիծ, այն իմաստով, որ թեստային դիտարկման դասի պիտակի որոշման վրա ազդեցություն ունեն միայն հարևան ուսուցողական դիտարկումները:

10 ՀՎ մեթոդի կիրառություն

Խնդրի դրվածքը. Որպեսզի նվազեցնի իր կորուստները, բանկը պետք է որոշում կայացնի՝ ում տրամադրել վարկ: Դիմորդի ժողովրդագրական և սոցիալ-տնտեսական իրավիճակը նախօրոք ստուգված է վարկային կառավարչի կողմից:

Խնդրի համար վերցրել ենք գերմանական բանկի վարկային տվյալներ, որի տվյալների նկարագրությունը որպես հղում տեղադրված է հավելվածում: Այդ տվյալները պարունակում են 20 փոփոխականներ և դիմորդների դասակարգում, արդյոք նա համարվում է լավ թե վատ վարկային ռիսկ կրող: 0-ն համարվում է վատ, իսկ 1-ը՝ լավ: Ներկայացված է 1000 հայտ:

Ակնկալվում է, որ մեր ունեցած տվյալների հիման վրա մշակված մոդելով կառավարիչը որոշում կայացնի. տրամադրել վարկ տվյալ դիմորդին, թե՝ ոչ:

Օգտվելով R վիճակագրական փաթեթից՝ դիմորդների մասին արել ենք հետևյալ եզրակացությունը:

0 1

300 700

այսինքն 70

Call:

svm(formula = Creditability ., data = Train, cost = 1, gamma = 1/length(Train),

probability = TRUE)

Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

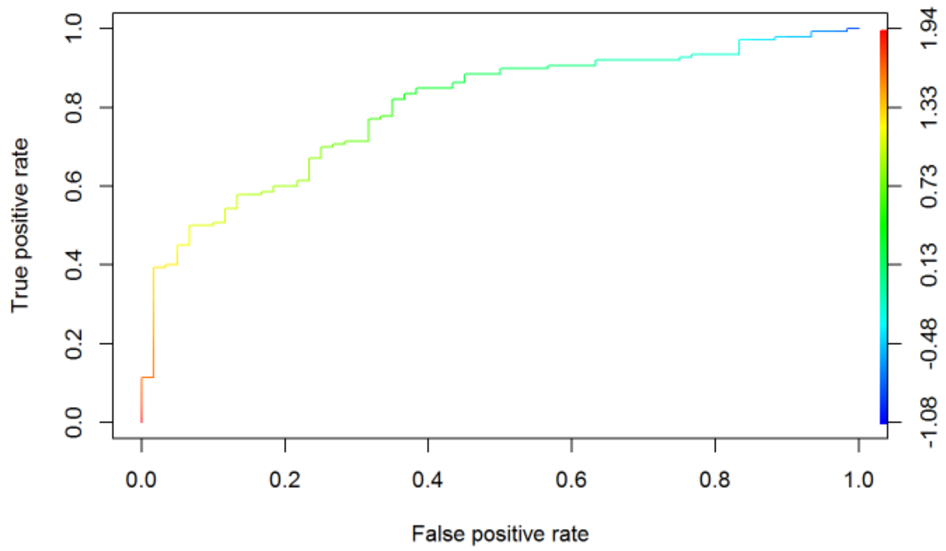
cost: 1

gamma: 0.04761905

Number of Support Vectors: 520
(281 239)
Number of Classes: 2
Levels:
0 1
Իսկ թեստ անլուց հետո հետևյալը`
Confusion Matrix and Statistics
Reference
Prediction 0 1
0 22 13
1 38 127
Accuracy : 0.745
95
No Information Rate : 0.7
P-Value [Acc > NIR] : 0.0934362
Kappa : 0.3108
Mcnemar's Test P-Value : 0.0007775
Sensitivity : 0.9071
Specificity : 0.3667
Pos Pred Value : 0.7697
Neg Pred Value : 0.6286
Prevalence : 0.7000
Detection Rate : 0.6350
Detection Prevalence : 0.8250
Balanced Accuracy : 0.6369
'Positive' Class : 1

Արդյունքում ստացվեց, որ Sensitivity=0.9071, ինչը նշանակում է լավ ռիսկ համարվող դիմորդները դասակարգչի կողմից համարվել են ճշտգրիտ 90.71

Կառուցենք ՀՎ մեթոդի արդյունքը Roc Curve and AUC(area under the curve)-ի միջոցով:



Վերջնական ունեցանք մոդելի ճշտգրտությունը՝ 0.7988095։

11 Հավելված

(Նկ.1)-ի օրինակը կառուցելու համար օգտագործվել է MATLAB ծրագիրը: function createfigure(X1, Y1, X2, Y2, X3, Y3, xdata1, ydata1, zdata1)

figure1 = figure;

axes1 = axes('Parent',figure1); hold(axes1,'on');

plot(X1,Y1,'DisplayName','setosa','Marker','+','LineStyle','none',... 'Color',[1 0 0]);

plot(X2,Y2,'DisplayName','versicolor','Marker','*','LineStyle','none',... 'Color',[0 1 0]);

plot(X3,Y3,'DisplayName','Support Vectors','Marker','o','LineStyle','none',... 'Color',[0 0 0]);

contour(xdata1,ydata1,zdata1,'LineColor',[0 0 0],'LevelList',0);

box(axes1,'on'); axis(axes1,'tight'); legend(axes1,'show');

ՀՎ մեթոդի կիրառության համար R վիճակագրական փաթեթը SVM library(caret) Loading required package: lattice Loading required package: ggplot2 library(gplots) Attaching package: 'gplots' The following object is masked from 'package:stats': lowess library(ROCR) library(caTools) Warning: package 'caTools' was built under R version 3.3.3 library(ggplot2) library("lattice") library(e1071) data<-read.csv("index.csv") str(data) 'data.frame': 1000 obs. of 21 variables: $Creditability : int1111111111...$ Account.Balance : int 1 1 2 1 1 1 1 1 4 2 ... $Duration.of.Credit..month. : int18912121210861824...$ Payment.Status.of.Previous.Credit: int 4 4 2 4 4 4 4 4 4 2 ... $Purpose : int2090000033...$ Credit.Amount : int 1049 2799 841 2122 2171 2241 3398 1361 1098 3758 ... $Value.Savings.Stocks : int1121111113...$ Length.of.current.employment : int 2 3 4 3 3 2 4 2 1 1 ... $Instalment.per.cent : int4223411241...$ Sex...Marital.Status : int 2 3 2 3 3 3 3 3 2 2 ... $Guarantors : int1111111111...$ Duration.in.Current.address : int 4 2 4 2 4 3 4 4 4 4 ... $Most.valuable.available.asset : int2111211134...$ Age..years. : int 21 36 23 39 38 48 39 40 65 23 ... $Concurrent.Credits : int3333133333...$ Type.of.apartment : int 1 1 1 1 2 1 2 2 2 1 ... $No.of.Credits.at.this.Bank : int1212222121...$ Occupation : int 3 3 2 2 2 2 2 2 1 1 ... $No.of.dependents : int1212121211...$ Telephone : int 1 1 1 1 1 1 1 1 1 1 ... $Foreign.Worker : int1112222211...data$Creditability<-as.factor(data$Creditability)data$Account.Balance<-as.factor(data$Account.Balance)data$ as.factor(data$Payment.Status.of.Previous.Credit)data$Purpose<-as.factor(data$Purpose$ as.factor(data$Value.Savings.Stocks)data$Length.of.current.employment<-

as.factor(data*Length.of.current.employment)data*Instalment.per.cent<-
as.factor(data*Instalment.per.cent)data*Sex...Marital.Status<-as.factor(data*Sex...Mari*
as.factor(data*Guarantors)data*Duration.in.Current.address<-as.factor(data*Duration*
as.factor(data*Duration.of.Credit..month.)data*Most.valuable.available.asset<-
as.factor(data*Most.valuable.available.asset)data*Concurrent.Credits<-
as.factor(data*Concurrent.Credits)data*Type.of.apartment<-as.factor(data*Type.of.apar*
as.factor(data*No.of.Credits.at.this.Bank)data*Occupation<-as.factor(data*Occupation)d*
as.factor(data*No.of.dependents)data*Telephone<-as.factor(data*Telephone)data*Foreign
as.factor(data*Foreign.Worker)summary(data*Creditability) 0 1 300
700 42.85

700/1000 [1] 0.7 set.seed(1982) trainIndex <- createDataPartition(data*Creditab*
.8, list = FALSE)Train < −data[trainIndex,]Test < −data[−trainIndex,]credit.svm <
−svm(Creditability ., data = Train, cost = 1, gamma = 1/length(Train), probability =
TRUE)summary(credit.svm)Call : svm(formula = Creditability ., data =
Train, cost = 1, gamma = 1/length(Train), probability = TRUE)Parameters :
SVM − Type : C − classificationSVM − Kernel : radialcost : 1gamma :
0.04761905NumberofSupportVectors : 520(281239)NumberofClasses : 2Levels :
01library(caret)Predict$_s$vm < −predict(credit.svm, newdata = Test, type =
*"prob")confusionMatrix(Predict$_s$vm, Test*Creditability,positive = "1")
Confusion Matrix and Statistics Reference Prediction 0 1 0
22 13 1 38 127 Accuracy : 0.745 95 No Information Rate :
0.7 P-Value [Acc > NIR] : 0.0934362 Kappa : 0.3108 Mcnemar's Test P-Value : 0.0007775 Sensitivity : 0.9071 Specificity :
0.3667 Pos Pred Value : 0.7697 Neg Pred Value : 0.6286 Prevalence : 0.7000 Detection Rate : 0.6350 Detection Prevalence
: 0.8250 Balanced Accuracy : 0.6369 'Positive' Class : 1 Roc
Curve and AUC(area under the curve) predicting the test data
svmmodel.predict<-predict(credit.svm,subset(Test,select=-Creditability),decision
svmmodel.probs<-attr(svmmodel.predict,"decision.values") svmmodel.class<-
predict(credit.svm,Test,type="class") svmmodel.labels<-Test*Creditability*

roc analysis for test data svmmodel.prediction<-prediction(svmmodel.probs,svi
svmmodel.performance<-performance(svmmodel.prediction,"tpr","fpr")
svmmodel.auc<-performance(svmmodel.prediction,"auc")@y.values[[1]]
plot(svmmodel.performance, colorize=T) AUC-area under the curve

svmmodel.auc [1] 0.7988095

12 Գրականություն

- GarethJames,DanielaWitten,TrevorHastie, RobertTibshirani ”An Introductionton Statistical Learning with Applications in R”, Springer Science + Business Media, New York, 2013
- https://onlinecourses.science.psu.edu/stat857/node/222/
- https://onlinecourses.science.psu.edu/stat857/node/215/
- http://detexify.kirelabs.org/classify.html
- https://en.wikipedia.org/wiki/Support$_{vector_machine}$