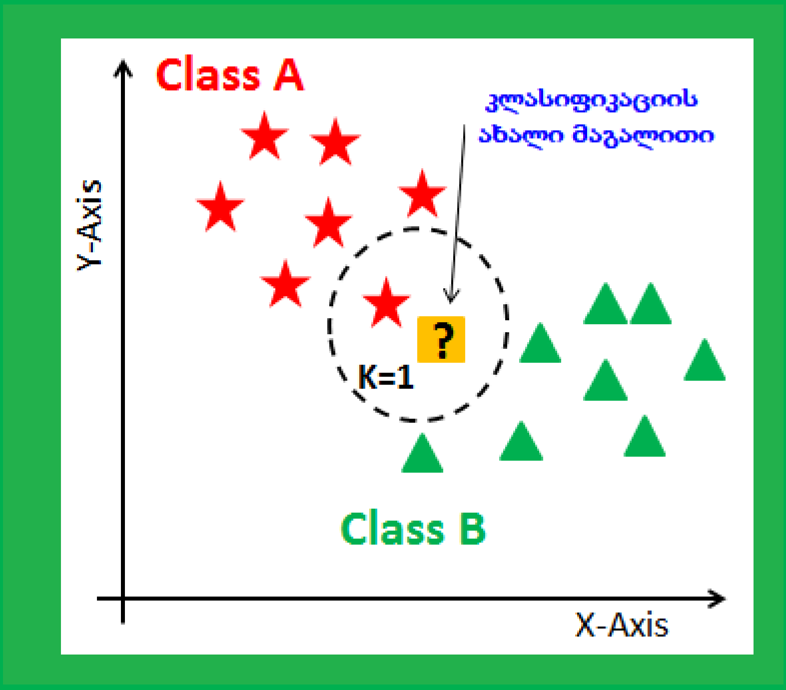
დოქტორანტის ბიბლიოთეკა

ქართლოს ყაჭიაშვილი

მანქანური სწავლების მეთოდები და ალგორითმები

(სემინარის მეთოდმითითებანი)



„სტუ-ს IT- კონსალტინგის სამეცნიერო ცენტრი"

საქართველოს ტექნიკური უნივერსიტეტი

ქართლოს ყაჭიაშვილი

მანქანური სწავლების მეთოდები და ალგორითმები

(სემინარული სამუშაოს მეთოდური მითითებანი)

დამტკიცებულია:

სტუ-ს „IT კონსალტინგის სამეცნიერო ცენტრის“ სარედაქციო კოლეგიის მიერ ოქმი N2, 10.02.2021

თბილისი

2021

უაკ 004.5

განხილულია მანქანური სწავლების ძირითადი მეთოდები, მათი რეალიზაციის ალგორითმები და გამოყენების მაგალითები. კერძოდ, ისეთი საკითხები, როგორიცაა დამოკიდებულების აღდგენა, გადაწყვეტილების თეორია, ინფორმაციის თეორია, ალბათური განაწილებები, ალბათურ მოდელები, რეგრესიის წრფივი მოდელები, ბაიესური მოდელები, კლასიფიკაციის წრფივი მოდელები, მანქანური სწავლების ალგორითმები, ნეირონული ქსელები და ბირთვული მეთოდები. მეთოდური მითითებებში შემოთავაზებულია აღნიშნული თეორიული საკითხების განხილვა და შესაბამისი პრაქტიკული ამოცანების გადაწყვეტა დოქტორანტებთან ერთად სემინარულ მეცადინეობებზე. მეთოდური მითითებები რეკომენდებულია ინფორმატიკის სპეციალობის დოქტორანტებისათვის.

რეცენზენტი:

პროფ. ა. ფრანგიშვილი (სტუ, ტ.მ.დ.)

ISBN 978-9941-8-1753-3

სასწავლოკურსისმიზანი

შეასწავლოს სტუდენტებს მანქანური სწავლების ძირითადი მეთოდები და რეალიზაციის ალგორითმები, მათი გამოყენება პრაქტიკული ამოცანების გადასაწყვეტად; მისცეს სტუდენტებს კვლევის ობიექტის საწყის მონაცემთა შეგროვების, შენახვისა და შემდგომი ანალიზისათვის მომზადების ხერხები და ინსტრუმენტები, მათი ინტელექტუალური ანალიზის საფუძველზე ცოდნის გენერირების საშუალებები.

საგნის შესწავლის შედეგად მიღებული ცოდნა და შეძენილი უნარები

მანქანური სწავლების თანამედროვე მეთოდებისა და შესაბამისი პროგრამული უზრუნველყოფის გამოყენებით **განსაზღვრავს** საკვლევი ობიექტის პრობლემებს და მათი გადაწყვეტის ამოცანებს.

ინფორმაციის მოპოვების, შენახვისა და დამუშავების პრაქტიკული ამოცანების გადასაწყვეტად **განსაზღვრავს** მანქანური სწავლების მათემატიკური მეთოდების გამოყენების შესაძლებლობას.

მონაცემების ანალიზისა და მათი შემდგომი დამუშავების საფუძველზე **აკეთებს** დასაბუთებულ დასკვნებს.

მონაცეთა ინტელექტუალური ანალიზის საფუძველზე **ახდენს** ცოდნის გენერირების პროცესის ალგორითმების შემუშავებას.

დამოუკიდებლად **აწარმოებს** წარმოქმნილი პრობლემის გადაჭრას, შესაბამისი ინფორმაციის მოპოვებითა და ახალი ცოდნის გენერირებით.

ახდენს საკუთარი შეხედულების არგუმენტირებულად ფორმულირებას და ზეპირი ან წერილობითი ფორმით გადმოცემას. ასევე სამეცნიერო დისკუსიების წარმართვას.

საათების განაწილება (სტუდენტის დატვირთვა)

კრედიტების რაოდენობა 6. ლექცია - 30 სთ., სემინარი (ჯგუფში მუშაობა) –30 სთ. დამოუკიდებელი მუშაობა 137 სთ.

სემინარული მუშაობის თემები:

1. **შესავალი მანქანურ სწავლებაში.** გამოყენებითი მაგალითების განხილვა.

2. **ალბათობის თეორია.** ალბათობის სიმკვრივის კონკრეტული მაგალითების განხილვა. მათემატიკური მოლოდინის, კოვარიაციის გამოთვლა. ბაიესის ალბათობებზე მაგალითების განხილვა.

3. **დამოკიდებულების აღდგენა.** ბაიესის იდეის გამოყენებით დამოკიდებულების აღდგენის მაგალითების განხილვა. მოდელების შერჩევის მაგალითები.

4. **გადაწყვეტილების თეორია.** არასწორე კლასიფიკაციის დონის მინიმიზაციის მაგალითები. მოსალოდნელი დანაკარგების მინიმიზაციის მაგალითები. კონკრეტული დანაკარგების ფუნქციების განხილვა რეგრესიისათვის.

5. **ინფორმაციის თეორია.** შედარებითი ენტროპია და ურთიერთ ინფორმაცია. კონკრეტული მაგალითები გამოთვლა.

6. **ალბათური განაწილებები.** ბინარული ცვლადი. ბეტა და დირიხლეს განაწილებების კონკრეტული შემთხვევების განხილვა. მრავალგანზომილებიანი ცვლადების მაგალითები.

ალბათურ გენერირებადი მოდელების მაგალითები. ლოგისტიკური რეგრესიის მაგალითები. მრავალკლასიანი ლოგისტიკური რეგრესიის მაგალითების განხილვა.

7. **გაუსის განაწილება.** პირობითი გაუსის განაწილებების მაგალითების განხილვა. ზღვრული გაუსის განაწილებების შემთხვევების განხილვა. მაქსიმალური დამაჯერებლობის მეთოდით გაუსის განაწილების პარამეტრების შეფასებები. მიმდევრობითი მეთოდით პარამეტრების შეფასებების მონახვა.

8. **ექსპონენციალური განაწილებები.** სტიუდენტის განაწილება პრაქტიკული შესწავლა. ექსპონენციალური ოჯახი სხვადასხვა წარმომადგენლების განხილვა. არაპარამეტრული მეთოდების პრაქტიკული შესწავლა.

9. **რეგრესიის წრფივი მოდელები.** უდიდესი დამაჯერებლობის და უმცირეს კვადრატთა მეთოდების პრაქტიკული გამოყენება და უმცირეს კვადრატთა გეომეტრიის პრაქტიკლი განხილვა.

10. **ბაიესური წრფივი რეგრესია.** პარამეტრული განაწილებზების განხილვა. პროგნოზირებადი განაწილების მაგალითები. ექვივალენტური ბირთვების მაგალითების განხილვა.

11. **ბაიესური მოდელის შედარება.** თვალსაჩინოების ფუნქციის შეფასების კონკრეტული მაგალითები. თვალსაჩინოების ფუნქციის მაქსიმიზაციის მაგალითები. პარამეტრების ეფექტური რაოდენობის შერჩევის მაგალითები.

12. **კლასიფიკაციის წრფივი მოდელები.** დისკრიმინანტის ფუნქციები ორი კლასის შემთხვევაში და კლასიფიკაციის პრაქტიკული განხორციელება. უმცირეს კვადრატთა მეთოდის გამოყენება კლასიფიკაციისათვის. მრავლობითი კლასიფიკაცია ფიშერის დისკრიმინანტის გამოყენებით.

13. **მანქანური სწავლების ალგორითმები.** პრაქტიკული ამოცანების განხილვა. მონაცემთა ინტელექტუალური ანალიზი. Data Mining და დიდ მონაცემთა სტატისტიკური დამუშავების მაგალითების განხილვა

14. **ნეირონული ქსელები.** ქსელური სწავლების მაგალითები. პარამეტრული ოპტიმიზაციის მაგალითები. გრადიენტის ინფორმაციის გამოყენება დამოკიდებულების აპროქსიმაციისათვის.

15. **ბირთვული მეთოდები.** ბირთვების აგების მაგალითები. გაუსის პროცესების მაგალითები. რეგრესიული დამოკიდებულების აღდგენა გაუსის პროცესებში.

სემინარული სამუშაო # 3

სემინარის თემა: დამოკიდებულების აღდგენა.

განხილული იქნება შემთხვევით სიდიდეებს შორის დამოკიდებულების აღდგენის იდეა და მისი გამოყენება პრაქტიკული პრობლემების გადასაწყვეტად. კერძოდ, მაგალითის სახით, განვიხილავთ თუ როგორ შეიძლება წარმოვადგინოთ პოლინომიალური მრუდის მორგების პრობლემა შეცდომის მინიმიზაციის ტერმინებში. ასევე, მრუდის მორგების პრობლემა განხილული იქნება ალბათური პერსპექტივიდან, რისთვისაც შემოტანილი იქნება შეცდომების ფუნქციებისა და რეგულარიზაციის ცნებები, რომალსაც მივყავართ პრობლემის სრულ ბეიესურ გაგებასთან. სემინარზე განხილული საკითხები დაკავშირებულია კურსის სახელმძღვანელოს „მანქანური სწავლება, სახელმძღვანელო (ლექციების კურსი) უნივერსიტეტის სტუდენტებისათვის“, მომზადებული ქ.ი. ყაჭიაშვილის მიერ წიგნის „M. Bishop, Pattern

Recognition and Machine Learning. Springer Verlag“ გარკვეული პუნქტების თარგმანი, თავი 3-ში მოცემულ თეორიულ მასალასთან.

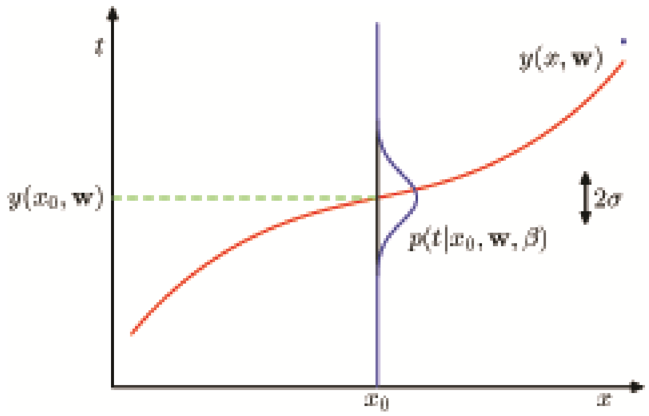
მრუდის მორგების პრობლემის მიზანია სამიზნე ცვლადის t -ს მნიშვნელობის პროგნოზირება შესაყვანი (დამოუკიდებელი) x ცვლადის ახალი მნიშვნელობისათვის სასწავლო მონაცემების სიმრავლის საფუძველზე, რომელიც მოიცავს შესაყვანი ცვლადის N მნიშვნელობებს $\mathbf{X}=(x_1,\ldots,x_N)^T$ და მათ შესაბამის სამიზნე ცვლადის მნიშვნელობებს $\mathbf{t}=(t_1,\ldots,t_N)^T$. გაურკვევლობა (განუსაზღვრელობა) სამიზნე ცვლადის მნიშვნელობის შესახებ შეიძლება გამოვხატოთ ალბათობის განაწილების გამოყენებით. ამ მიზნით, დავუშვათ, რომ x–ის მოცემული მნიშვნელობისათვის, t -ს შესაბამის მნიშვნელობას აქვს გაუსის განაწილებით

$$N\left(x \mid \mu,\sigma^2\right)=\frac{1}{\left(2\pi\sigma^2\right)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{2.42}$$

მოცემული პოლინომური მრუდით განსაზღვრული საშუალოთი $y(x,\mathbf{w})$. ამრიგად, გვაქვს

$$p\left(t \mid x,\mathbf{w},\beta\right)=N\left(t \mid y(x,\mathbf{w}),\beta^{-1}\right), \tag{3.1}$$

სადაც შემოვიტანეთ სიზუსტის პარამეტრი β, რომელიც შეესაბამება განაწილების შებრუნებულ დისპერსიას. ეს სქემატურად ილუსტრირებულია ნახაზ 3.1–ზე.



ნახ. 3.1. x–თვის t–ს გაუსის პირობითი განაწილების სქემატური ილუსტრაცია მოცემული (3.1)-ით, რომელშიც საშუალო მოცემულია პოლინომიალური ფუნქციით $y(x,\mathbf{w})$ და სიზუსტე მოცემულია β პარამეტრით, რომელიც დისპერსიასთან დაკავშირებულია დამოკიდებულებით $\beta^{-1}=\sigma^2$.

გამოვიყენოთ სასწავლო მონაცემები $\{\mathbf{X},\mathbf{t}\}$ უცნობი პარამეტრების $\mathbf{w}$–ს და β–ს მნიშვნელობების დასადგენად *მაქსიმალური დასაჯერობით.* თუ ვუშვებთ, რომ მონაცემები მიღებული (3.1)–დან არიან დამოუკიდებელი, მაშინ *დასაჯერობის ფუნქცია* მოიცემა ასე

$$p(\mathbf{t}\mid \mathbf{x},\mathbf{w},\beta)=\prod_{n=1}^{N}N\left(t_n\mid y(x_n,\mathbf{w}),\beta^{-1}\right). \quad (3.2)$$

ცნობილია, რომ მოსახერხებელია ალბათობის ფუნქციის ლოგარითმის მაქსიმიზაცია. (2.42)–ით მოცემული გაუსის განაწილების ფორმულის გამოყენებით, ვიღებთ დასაჯერობის ფუნქციის ლოგარითმს შემდეგი სახით

$$\ln p(\mathbf{t}\mid \mathbf{x},w,\beta)=-\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n,w)-t_n\}^2+\frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi). \quad (3.3)$$

პირველ რიგში განვიხილოთ მაქსიმალური დასაჯერობის ამოხსნით პოლინომური კოეფიციენტების განსაზღვრა, რომელსაც ავღნიშნავთ $\mathbf{w}_{ML}$. ეს განისაზღვრება (3.3)–ის $\mathbf{w}$–ით მაქსიმიზირებით. ამ მიზნით, შეგვიძლია გამოვტოვოთ (3.3)-ის მარჯვენა მხარის ბოლო ორი წევრი, რადგან ისინი არ არიან დამოკიდებული $\mathbf{w}$–ზე. ასევე, შევნიშნოთ, რომ ალბათობის ლოგარითმის დადებითი მუდმივი კოეფიციენტით მასშტაბირება არ ცვლის მაქსიმუმის მდებარეობას $\mathbf{w}$–ს მიმართ და, ამიტომ, შეგვიძლია ჩავანაცვლოთ $\beta/2$ კოეფიციენტი $1/2$–ით. ბოლოს, დასაჯერობის ლოგარითმის მაქსიმიზაციის ნაცვლად, შეგვიძლია ექვივალენტურად დასაჯერობის ლოგარითმის უარყოფითი მნიშვნელობის მინიმიზაცია. ამიტომ ვხედავთ, რომ, როდესაც საქმე ეხება $\mathbf{w}$–ს განსაზღვრას, დასაჯერობის ლოგარითმის მაქსიმიზაცია ექვივალენტურია *შეცდომის კვადრატების ჯამის ფუნქციის* მინიმიზაციის. ამრიგად, შეცდომის კვადრატების ჯამის ფუნქცია წარმოიშვა დასაჯერობის მაქსიმიზაციის შედეგად იმ დაშვებით, რომ ხმაურს აქვს გაუსის განაწილება.

ასევე შეგვიძლია გამოვიყენოთ მაქსიმალური დასაჯერობა გაუსის პირობითი განაწილების β სიზუსტის პარამეტრის დასადგენად. (3.3)–ის მაქსიმიზაცია β–ს მიმართ გვაძლევს

$$\frac{1}{\beta_{ML}}=\frac{1}{N}\sum_{n=1}^{N}\{y(x_n,w_{ML})-t_n\}^2. \quad (3.4)$$

ამრიგად პირველად უნდა განვსაზღვროთ პარამეტრ ვექტორი $\mathbf{w}_{ML}$, რომელიც მართავს საშუალო მნიშვნელობას და შემდეგ გამოვიყენოთ ის, რომ ვიპოვოთ β_{ML} სიზუსტის პარამეტრი.

$\mathbf{w}$ და β პარამეტრების დადგენის შემდეგ, შეგვიძლია გავაკეთოთ პროგნოზები x–ის ახალი მნიშვნელობებისთვის. რადგან, ახლა გვაქვს ალბათური მოდელი, ისინი არიან წარმოდგენილი *საპროგნოზო განაწილების* საშუალებით, რომელიც გვაძლევს t–ს ალბათობის განაწილებას, და არა უბრალოდ წერტილოვან შეფასებას და მიიღება მაქსიმალური დასაჯერობით მიღებული პარამეტრების ჩასმით (3.1)–ში, რაც გვაძლევს

$$p(t\mid x,\mathbf{w}_{ML},\beta_{ML})=N\left(t\mid y(x,w_{ML}),\beta_{ML}^{-1}\right). \quad (3.5)$$

ეხლა დავიწყოთ ბაიესური მიდგომის განხილვა და, ამისათვის, შემოვიტანოთ აპრიორული განაწილება მრავალწევრის $\mathbf{w}$ კოეფიციენტებზე. სიმარტივისათვის, განვიხილოთ შემდეგი ფორმის გაუსის განაწილება

$$p(\mathbf{w}|\alpha)=N(\mathbf{w}|\mathbf{0},\alpha^{-1}\mathbf{I})=\left(\frac{\alpha}{2\pi}\right)^{(M+1)/2}\exp\left\{-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right\}, \qquad (3.6)$$

სადაც α არის განაწილების სიზუსტე, და $M+1$ არის $\mathbf{w}$ ვექტორის ელემენტების საერთო რაოდენობა M–ური რიგის პოლინომისათვის. ისეთ ცვლადებს, როგორიცაა α, რომლებიც აკონტროლებენ მოდელის პარამეტრების განაწილებას, *ჰიპერპარამეტრებს* უწოდებენ. ბაიესის თეორემის გამოყენებით, $\mathbf{w}$–ს აპოსტერიორული განაწილება პროპორციულია აპრიორული განაწილების და დასაჯერობის ფუნქციის ნამრავლის

$$p(\mathbf{w}|\mathbf{x},\mathbf{t},\alpha,\beta)\propto p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)p(\mathbf{w}|\alpha). \qquad (3.7)$$

ეხლა შეგვიძლია განვსაზღვროთ $\mathbf{w}$ მონაცემების გათვალისწინებით მისი ($\mathbf{w}$-ს) ყველაზე ალბათური მნიშვნელობის მოძებვნით, სხვა სიტყვებით რომ ვთქვათ, აპოსტერიორული განაწილების მაქსიმიზაციით. ამ ტექნიკას ეწოდება *მაქსიმალური აპოსტერიორი*, ან უბრალოდ მაპ (MAP). (3.7)-ის უარყოფითი ლოგარითმის აღებით და (3.3)-თან და (3.6)-თან გაერთიანებით ვიპოვით, რომ აპოსტერიორის მაქსიმუმი მოცემულია შემდეგი გამოსახულების მინიმიზაციით

$$\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2+\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}. \qquad (3.8)$$

ამრიგად, ვხედავთ, რომ აპოსტერიორული განაწილების მაქსიმიზაცია ექვივალენტურია რეგულარიზებული შეცდომის კვადრატების ჯამის ფუნქციის მინიმიზაციისა $\lambda=\alpha/\beta$ რეგულარიზაციის პარამეტრით.

თუმცა, ჩვენ გამოვიყენეთ აპრიორული განაწილება $p(\mathbf{w}|\alpha)$, მაგრამ ჯერჯერობით მაინც შორსა ვართ $\mathbf{w}$–ს წერტილოვანი შეფასების პოვნისაგან და ამდენად ეს ჯერ კიდევ არ წარმოადგენს ბაიესურ დამუშავებას. სრული ბაიესური მიდგომისათვის, თანმიმდევრულად უნდა გამოვიყენოთ ალბათობის ჯამის და გამრავლების წესები, რაც მოითხოვს, როგორც მალე ვნახავთ, $\mathbf{w}$–ს ყველა მნიშვნელობებზე ინტეგრალის აღებას. ასეთი მარგინალიზაცია ჯდება სახეთა გამოცნობის ბაიესური მეთოდების გულში.

მრუდის მორგების პრობლემაში მოცემულია სასწავლო მონაცემები $\mathbf{X}$ და $\mathbf{t}$, ახალ x სატესტო წერტილთან ერთად და ჩვენი მიზანია t–ს მნიშვნელობის პროგნოზირება. ამიტომ გვინდა შევაფასოთ პროგნოზული განაწილება $p(t|x,\mathbf{X},\mathbf{t})$. აქ ვგულისხმობთ, რომ α და β პარამეტრები არიან დაფიქსირებული და წინასწარ ცნობილი.

ბაიესური დამუშავება უბრალოდ შეესაბამება ალბათობის შეკრებისა და გამრავლების წესების თანმიმდევრულ გამოყენებას, რაც საშუალებას იძლევა საპროგნოზო განაწილება ჩაიწეროს ასე

$$p(t \mid x, \mathbf{X}, \mathbf{t}) = \int p(t \mid x, \mathbf{w}) p(\mathbf{w} \mid \mathbf{X}, \mathbf{t}) d\mathbf{w}. \qquad (3.9)$$

აქ $p(t \mid x, \mathbf{w})$ მოცემულია (3.1)-ით და ჩვენ გამოვტოვეთ დამოკიდებულება α და β-ზე აღნიშვნების გასამარტივებლად. აქ $p(\mathbf{w} \mid \mathbf{X}, \mathbf{t})$ არის პარამეტრების აპოსტერიორული განაწილება და მისი პოვნა შესაძლებელია (3.7)–ის მარჯვენა მხარის ნორმალიზებით. ისეთი პრობლემებისათვის, როგორიცაა მრუდის მორგების მაგალითი, ეს აპოსტერიორული განაწილება არის გაუსის ტიპის და შეიძლება შეფასდეს ანალიტიკურად. ანალოგიურად, (3.9)-ში ინტეგრება ასევე შეიძლება განხორციელდეს ანალიტიკურად იმ შედეგით, რომ საპროგნოზო განაწილებას აქვს შემდეგი სახის გაუსის განაწილება

$$p(t \mid x, \mathbf{x}, \mathbf{t}) = N\left(t \mid m(x), s^2(x)\right), \qquad (3.10)$$

სადაც საშუალო და დისპერსია არიან მოცემული ფორმულებით

$$m(x) = \beta \phi(x)^T \mathbf{S} \sum_{n=1}^{N} \phi(x_n) t_n, \qquad (3.11)$$

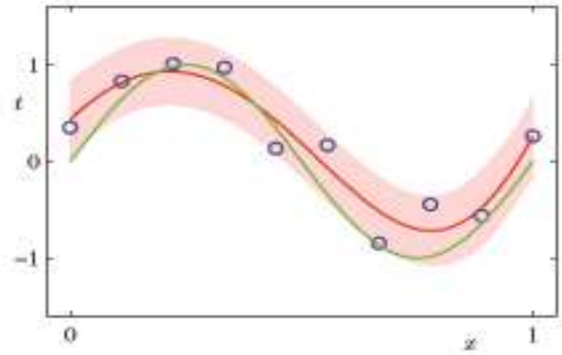
$$s^2(x) = \beta^{-1} + \phi(x)^T \mathbf{S} \phi(x). \qquad (3.12)$$

აქ მატრიცა $\mathbf{S}$ მოცემულია ფორმულით

$$\mathbf{S}^{-1} = \alpha \mathbf{I} + \beta \sum_{n=1}^{N} \phi(x_n) \phi(x)^T, \qquad (3.13)$$

სადაც $\mathbf{I}$ არის ერთეულოვანი მატრიცა და ვექტორი $\phi(x)$ განსაზღვრულია ელემენტებით $\phi_i(x) = x^i$, $i = 0, \ldots, M$ -თვის.

ვხედავთ, რომ (3.10)–ის საპროგნოზო განაწილების დისპერსია, ისევე როგორც საშუალო, დამოკიდებულია x–ზე. (3.12)–ის პირველი წევრი წარმოადგენს გაურკვევლობას t-ს საპროგნოზო მნიშვნელობაში სამიზნე ცვლადებში ხმაურის გამო და უკვე იყო გამოხატულია მაქსიმალური დასაჯერობის პროგნოზულ განაწილებაში (3.5)–ში β_{ML}^{-1}-ით. ამასთან, მეორე წევრი წარმოიქმნება $\mathbf{w}$ პარამეტრების გაურკვევლობით და ეს არის ბაიესური დამუშავების შედეგი. სინთეზური სინუსოიდული რეგრესიის პრობლემისათვის საპროგნოზო განაწილება მოცემულია ნახაზ 3.2–ზე.



ნახ. 3.2. საპროგნოზო განაწილება, რომელიც წარმოიქმნება პოლინომიალური მრუდის ბაიესური დამუშავების შედეგად, $M=9$ მრავალწევრის გამოყენებით, ფიქსირებული პარამეტრებით $\alpha=5\times10^{-3}$ და $\beta=11.1$ (შეესაბამება ხმაურის ცნობილ დისპერსიას), რომელშიც წითელი მრუდი აღნიშნავს საპროგნოზო განაწილების საშუალო მნიშვნელობას და წითელი არე შეესაბამება ±1 სტანდარტულ გადახრას საშუალოს გარშემო.

თეორიულ საკითხებზე მუშაობის შემდეგ განხილული იქნება პრაქტიკული მაგალითები სემინარის თემის მიხედვით და მათი ამოხსნის მეთოდები.

სემინარული სამუშაო # 4

სემინარის თემა: გადაწყვეტილების თეორია.

გადაწყვეტილების თეორია, ალბათობის თეორიასთან შერწყმისას, საშუალებას გვაძლევს მივიღოთ ოპტიმალური გადაწყვეტილებები სიტუაციებში, რომლებიც შეიცავენ გაურკვევლობას, ისეთებს, როგორებიც გვხვდება, მაგალითად, სახეთა გამოცნობაში. სემინარის მუშაობისას განხილული იქნება შემდეგი საკითხები: არასწორე კლასიფიკაციის დონის მინიმიზაცია; მოსალოდნელი დანაკარგების მინიმიზაცია; შეფასებისა და გადაწყვეტილების მიღების მეთოდები; დანაკარგების ფუნქცია რეგრესიისათვის.

გადაწყვეტილების თეორიის საგანია გვითხრას, თუ როგორ უნდა მივიღოთ ოპტიმალური გადაწყვეტილებები მოცემული შესაბამისი ალბათობით. განვიხილავთ გადაწყვეტილების მიღებისა და დასკვნის გაკეთების თეორიის მთავარ იდეებს, მათ თავისებურებებს და გამოყენების სპეციფიკას. *სემინარზე განხილული საკითხები დაკავშირებულია კურსის სახელმძღვანელოს „მანქანური სწავლება, სახელმძღვანელო (ლექციების კურსი) უნივერსიტეტის სტუდენტებისათვის“, მომზადებული ქ.ი. ყაჭიაშვილის მიერ წიგნის „M. Bishop, Pattern Recognition and Machine Learning. Springer Verlag“ გარკვეული პუნქტების თარგმანი, თავი 4-ში მოცემულ თეორიულ მასალასთან.*

დავუშვათ, რომ გვაქვს შეყვანის ვექტორი $\mathbf{x}$ შესაბამის $\mathbf{t}$ სამიზნე ცვლადების ვექტორთან ერთად და ჩვენი მიზანია ვიწინასწარმეტყველოთ $\mathbf{t}$–ს მნიშვნელობა, $\mathbf{x}$–ის ახალი მოცემული მნიშვნელობისათვის. რეგრესიის პრობლემებისათვის $\mathbf{t}$ მოიცავს უწყვეტ ცვლადებს, მაშინ როდესაც კლასიფიკაციის პრობლემებისთვის $\mathbf{t}$ წარმოადგენს კლასის ჭდეებს. ალბათობის ერთობლივი განაწილება $p(\mathbf{x},\mathbf{t})$ იძლევა ამ ცვლადებთან დაკავშირებულ გაურკვევლობაზე სრულ ინფორმაციას. $p(\mathbf{x},\mathbf{t})$–ს განსაზღვრა სასწავლო მონაცემების საფუძველზე არის დასკვნების მაგალითი და, როგორც წესი, ძალიან რთული პრობლემაა, რომლის გადაწყვეტას ეძღვნება ჩვენი კურსის დიდი ნაწილი. პრაქტიკული გამოყენებისას, ხშირად საჭიროა მოვახდინოთ $\mathbf{t}$-ს მნიშვნელობის პროგნოზირება, ან, უფრო ზოგადად, ჩვენი ცოდნის საფუძველზე იმასთან დაკავშირებით თუ რა მნიშვნელობებს ღებულობს $\mathbf{t}$, უნდა განვახორციელოთ $\mathbf{t}$–ს კონკრეტული პროგნოზი და ეს ასპექტი არის გადაწყვეტილების თეორიის საგანი.

განვიხილოთ, მაგალითად, სამედიცინო დიაგნოზის პრობლემა, რომელშიც ჩვენ გადავიღეთ პაციენტის რენტგენის სურათი და გვსურს დავადგინოთ, აქვს თუ არა პაციენტს კიბო. ამ შემთხვევაში, შემავალი ვექტორი $\mathbf{x}$ არის პიქსელების ინტენსივობის სიმრავლე სურათზე და გამომავალი ცვლადი t წარმოადგენს სიმსივნის არსებობას, რასაც ჩვენ აღვნიშნავთ C_1 კლასით, ან კიბოს არარსებობით, რომელსაც აღვნიშნავთ C_2 კლასით. მაგალითად, ჩვენ შეგვიძლია t ავირჩიოთ ორობითი ცვლადი სახით, ისე, რომ $t=0$ შეესაბამება C_1 კლასს და $t=1$ შეესაბამება C_2 კლასს. მოგვიანებით ვნახავთ, რომ ჭდის მნიშვნელობების არჩევა განსაკუთრებით მოსახერხებელია ალბათური მოდელებისათვის. ზოგადი დასკვნის პრობლემა შემდეგში გულისხმობს ერთობლივი განაწილების $p(\mathbf{x},C_k)$, ან ეკვივალენტურად $p(\mathbf{x},t)$–ს განსაზღვრას, რაც სიტუაციის ყველაზე სრულყოფილ ალბათურ აღწერას გვაძლევს. თუმცა, ეს შეიძლება იყოს ძალიან სასარგებლო და ინფორმაციული სიდიდე, მაგრამ საბოლოოდ უნდა გადავწყვიტოთ პაციენტს მკურნალობა ჩავუტაროთ თუ არა, და გვსურს, რომ ეს არჩევანი იყოს ოპტიმალური გარკვეული თვალსაზრისით. ეს არის *გადაწყვეტილების* ნაბიჯი და გადაწყვეტილების თეორიის საგანია გვითხრას, თუ როგორ უნდა მივიღოთ ოპტიმალური გადაწყვეტილებები მოცემული შესაბამისი ალბათობით. ჩვენ ვნახავთ, რომ გადაწყვეტილების ეტაპი ზოგადად ძალიან მარტივია, ტრივიალურიც კი, მას შემდეგ რაც გადავჭრით დასკვნის პრობლემას.

უფრო დეტალური ანალიზის დაწყებამდე, პირველ რიგში განვიხილოთ არაფორმალურად, თუ როგორი მოლოდინი შეიძლება გვქონდეს ალბათობის როლის შესახებ გადაწყვეტილების მიღებისას. როდესაც მივიღებთ რენტგენის სურათის $\mathbf{x}$ გამოსახულებას ახალი პაციენტისათვის, ჩვენი მიზანია გადავწყვიტოთ, რომელ კლასს მივაკუთვნოთ ეს გამოსახულება. ჩვენთვის საინტერესოა მოცემული სურათისათვის ორი კლასის ალბათობები, რომლებიც მოცემულია $p(C_k\,|\,\mathbf{x})$-ით. ბეიესის თეორემის გამოყენებით, ეს ალბათობები შეიძლება მოცემული იყოს შემდეგი სახით

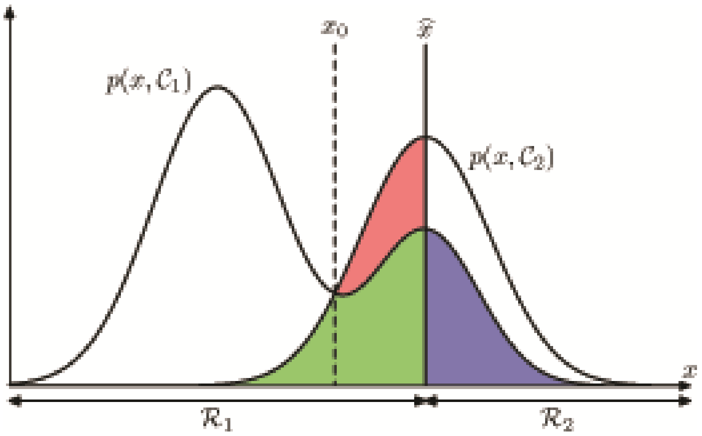
$$p(C_k\,|\,\mathbf{x})=\frac{p(\mathbf{x}\,|\,C_k)\,p(C_k)}{p(\mathbf{x})}. \qquad (4.1)$$

შევნიშნოთ, რომ ბეიესის თეორემაში მოცემული ნებისმიერი სიდიდე შეიძლება მივიღოთ ერთობლივი განაწილებიდან $p(\mathbf{x}, C_k)$ შესაბამისი ცვლადების ან მარგინალიზაციის ან განპირობების გზით. ახლა შეგვიძლია $p(C_k)$-ს ინტერპრეტაცია როგორც C_k კლასის აპრიორული ალბათობა, და $p(C_k \mid \mathbf{x})$ როგორც შესაბამისი აპოსტერიორული ალბათობა. ამრიგად, $p(C_1)$ წარმოადგენს ალბათობას, რომ ადამიანს აქვს კიბო, მანამ, სანამ რენტგენის გაზომვას გავაკეთებთ. ანალოგიურად, $p(C_1 \mid \mathbf{x})$ არის შესაბამისი ალბათობა, გადასინჯული ბაიესის თეორემის გამოყენებით, რენტგენის სურათში მოცემული ინფორმაციის გათვალისწინებით. თუ ჩვენი მიზანია შევამციროთ $\mathbf{x}$-ის არასწორი კლასისთვის მიკუთვნების შანსი, მაშინ ინტუიციურად უნდა ავირჩიოთ კლასი, რომელსაც აქვს უფრო მაღალი აპოსტერიორული ალბათობა. ახლა ჩვენ ვაჩვენებთ, რომ ეს ინტუიცია სწორია და ასევე განვიხილავთ გადაწყვეტილების მიღების უფრო ზოგად კრიტერიუმებს.

დავუშვათ, რომ ჩვენი მიზანია, უბრალოდ, რაც შეიძლება ნაკლები არასწორი კლასიფიკაცია გავაკეთოთ. ჩვენ გვჭირდება წესი, რომელიც $\mathbf{x}$–ის თითოეულ მნიშვნელობას მიანიჭებს ერთ რომელიმე კლასს. ასეთი წესი მონაცემების სივრცეს დაყოფს R_k არეებად, რომლებსაც ეწოდებათ *გადაწყვეტილების არეები*, ყოველი კლასისთვის თითო, ისეთი, რომ R_k-ს ყველა წერტილი მიეკუთვნება C_k კლასს. გადაწყვეტილების არეებს შორის საზღვრებს *გადაწყვეტილების საზღვრებს* ან *გადაწყვეტილების ზედაპირებს* უწოდებენ. შევნიშნოთ, რომ თითოეული გადაწყვეტილების არე არ არის აუცილებელი, რომ იყოს უწყვეტი, მაგრამ ის უნდა შეიცავდეს არაკვეთადი არეების გარკვეულ რაოდენობას. იმისათვის, რომ ვიპოვოთ გადაწყვეტილების ოპტიმალური წესი, უპირველეს ყოვლისა განვიხილოთ ორი კლასის შემთხვევა, მაგალითად, როგორც კიბოს პრობლემის შემთხვევაში. შეცდომა წარმოიქმნება, როდესაც მონაცემების ვექტორი, რომელიც მიეკუთვნება C_1 კლასს, მიეკუთვნება C_2 კლასს ან პირიქით. ამის ალბათობა მოიცემა ფორმულით

$$\begin{aligned} p(mistake) &= p(\mathbf{x} \in R_1, C_2) + p(\mathbf{x} \in R_2, C_1) = \\ &= \int_{R_1} p(\mathbf{x}, C_2)\, d\mathbf{x} + \int_{R_2} p(\mathbf{x}, C_1)\, d\mathbf{x}. \end{aligned} \tag{4.2}$$

ჩვენ თავისუფალი ვართ ავირჩიოთ გადაწყვეტილების წესი, რომელიც თითოეულ $\mathbf{x}$ წერტილს აკუთვნებს ორი კლასიდან ერთ–ერთს. ცხადია, $p(mistake)$–ის (p(*შეცდომის*)) შესამცირებლად უნდა ვეცადოთ, რომ თითოეული $\mathbf{x}$ მივაკუთვნოთ იმ კლასს, რომელსაც აქვს ინტეგრანდების მცირე მნიშვნელობები (4.2)–ში. ამრიგად, თუ $p(\mathbf{x}, C_1) > p(\mathbf{x}, C_2)$ $\mathbf{x}$–ის მოცემული მნიშვნელობისათვის, მაშინ ეს $\mathbf{x}$ უნდა მივაკუთვნოთ C_1 კლასს. ალბათობის გამრავლების წესიდან გვაქვს $p(\mathbf{x}, C_k) = p(C_k \mid \mathbf{x}) p(\mathbf{x})$. იმის გამო, რომ თანამამრავლი $p(\mathbf{x})$ საერთოა ორივე წევრისათვის, შეგვიძლია ახლიდან განვმარტოთ ეს შედეგი და ვთქვათ, რომ შეცდომის დაშვების მინიმალური ალბათობა მიიღება, თუ $\mathbf{x}$-ის ყოველი მნიშვნელობა ენიჭება იმ კლასს, რომლისთვისაც აპოსტერიორული ალბათობა $p(C_k \mid \mathbf{x})$ არის ყველაზე დიდი. ეს შედეგი ილუსტრირებულია ორი კლასისთვის და ერთგანზომილებიანი დაკვირვების x ცვლადისათვის, ნახაზ 4.1–ზე.



ნახ. 4.1. ორი კლასიდან თითოეულისათვის $p(\mathbf{x}, C_k)$ ერთობლივი ალბათობის სქემატური ილუსტრაცია, დახაზული x–თან მიმართებაში, გადაწყვეტილების $x = \hat{x}$ საზღვართან ერთად. მნიშვნელობები $x \geq \hat{x}$ კლასიფიცირდება როგორც C_2 კლასი და, შესაბამისად, მიეკუთვნება R_2 გადაწყვეტილების არეს, ხოლო წერტილები $x < \hat{x}$ კლასიფიცირდება როგორც C_1 და ეკუთვნის R_1–ს. შეცდომები წარმოიქმნება ლურჯი, მწვანე და წითელი არეებიდან, ასე რომ $x < \hat{x}$–თვის შეცდომები არიან იმ წერტიელბის გამო C_2 კლასიდან რომლებიც არასწორედ არიან კლასიფიციებული როგორც C_1–ი (წარმოდგენილია წითელი და მწვანე არეების ჯამით), და პირიქით, $x \geq \hat{x}$ არის წერტილებისთვის შეცდომები გამოწვეულია C_1 კლასიდან, რომლებიც არასწორად არიან კლასიფიცირებული როგორც C_2 (წარმოდგენილია ცისფერი არით). როდესაც ვცვლით გადაწყვეტილების საზღვრის $\hat{x}$–ის მდებარეობას, ლურჯი და მწვანე არეების გაერთიანება რჩება უცვლელი, ხოლო წითელი არის ზომა იცვლება. $\hat{x}$–ის ოპტიმალური არჩევანია ის, სადაც $p(\mathbf{x}, C_1)$ და $p(\mathbf{x}, C_2)$ მრუდები იკვეთება, რომელსაც შეესაბამება $\hat{x} = x_0$, რადგან ამ შემთხვევაში წითელი არე ქრება. ეს ექვივალენტურია გადაწყვეტილების წესისა მინიმალური არასწორი კლასიფიკაციის დონით, რომელიც x–ის ყოველ მნიშვნელობას ანიჭებს იმ კლასს, რომელსაც უფრო დიდი პოსტერიორული ალბათობა აქვს.

უფრო ზოგად შემთხვევაში, როდესავ კლასების რაოდენობა არის K, ოდნავ უფრო ადვილია მაქსიმიზაცია გავუკეთოთ სწორე გადაწყვეტილების ალბათობას, რომელიც მოიცემა ასე

$$p(correct) = \sum_{k=1}^{K} p(\mathbf{x} \in R_k, C_k) =$$

$$= \sum_{k=1}^{K} \int_{R_k} p(\mathbf{x}, C_k) d\mathbf{x} \qquad (4.3)$$

რაც მაქსიმალურად იზრდება როდესაც R_k არეები აირჩევა ისე, რომ თითოეული $\mathbf{x}$ მიეკუთვნება იმ კლასს, რომლისთვისაც $p(\mathbf{x}, C_k)$ არის უდიდესი. კვლავ, გამოვიყენოთ გამრავლების წესი $p(\mathbf{x}, C_k) = p(C_k \mid \mathbf{x}) p(\mathbf{x})$, და შევნიშნოთ, რომ მამრავლი $p(\mathbf{x})$ საერთოა ყველა წევრში.

დავინახავთ, რომ თითოეული $\mathbf{x}$ უნდა მიენიჭოს იმ კლასს, რომელსაც აქვს ყველაზე დიდი აპოსტერიორული ალბათობა $p(C_k\,|\,\mathbf{x})$.

თეორიულ საკითხებზე მუშაობის შემდეგ განხილული იქნება პრაქტიკული მაგალითები სემინარის თემის მიხედვით და მათი ამოხსნის მეთოდები.

სემინარული სამუშაო # 9

სემინარის თემა: რეგრესიის წრფივი მოდელები.

სემინარზე განხილული იქნება მასწავლებლით სწავლების პრობლემები რეგრესიული ანალიზის მაგალითზე. რეგრესიის მიზანი არის ერთი ან მეტი უწყვეტი სამიზნე t ცვლადების მნიშვნელობების პროგნოზირება *შესავალი* ცვლადების D-განზომილებიანი $\mathbf{x}$ ვექტორის მნიშვნელობის გათვალისწინებით. კურსის დასაწყისში განხილული რეგრესიის პრობლემის მაგალითის, როგორიცაა პოლინომიალური მრუდის მორგება, განვითარებისა და გაღრმავების მიზნით, განხილული იქნება რეგრესიული ანალიზის პრობლემა. პოლინომი არის კონკრეტული მაგალითი ფუნქციების ფართო კლასის, რომელსაც წრფივ რეგრესიულ მოდელებს უწოდებენ, რომლებიც არიან დასარეგულირებელი პარამეტრების წრფივი ფუნქციები და რომლებიც განხილული იქნებიან სემინარის მუშაობისას. კერძოდ, წრფივი რეგრესიული მოდელების უმარტივესი ფორმასთან, როდესაც რეგრესორები წარმოადგენენ შეყვანის (დამოუკიდებელი) ცვლადების წრფივ ფუნქციებს, განხილული იქნება ფუნქციების ბევრად უფრო სასარგებლო კლასი შეყვანის (დამოუკიდებელი) ცვლადების არაწრფივი ფუნქციების ფიქსირებული სიმრავლის წრფივი კომბინაციების გამოყენებით, რომლებიც ცნობილია როგორც *ბაზისური ფუნქციები.* ასეთი მოდელები წარმოადგენენ პარამეტრების წრფივ ფუნქციებს, რაც მათ აძლევს მარტივ ანალიტიკურ თვისებებს მაშინაც კი, როდესაც ასინი არიან არაწრფივი შეყვანის (დამოუკიდებელ) ცვლადებთან მიმართებაში. ასეთი მოდელებისათვის განხილული იქნება უცნობი პარამეტრების პოვნის უდიდესი დასაჯერობის და უმცირეს კვადრატთა მეთოდები. განხილული იქნება: უმცირეს კვადრატთა გეომეტრია; მიმდევრობითი სწავლება; რეგულირებადი უმცირეს კვადრატთა მეთოდი და მრავლობითი შედეგები. სემინარზე განხილული საკითხები დაკავშირებულია კურსის სახელმძღვანელოს „მანქანური სწავლება, სახელმძღვანელო (ლექციების კურსი) უნივერსიტეტის სტუდენტებისათვის“, მომზადებული ქ.ი. ყაჭიაშვილის მიერ წიგნის „M. Bishop, Pattern Recognition and Machine Learning. Springer Verlag“ გარკვეული პუნქტების თარგმანი, თავი 9-ში მოცემულ თეორიულ მასალასთან.

მოცემულია სასწავლო მონაცემთა სიმრავლე, რომელიც შეიცავს N დაკვირვებებს $\{\mathbf{x}_n\}$, სადაც $n=1,\ldots,N$, შესაბამის $\{t_n\}$ სამიზნე მნიშვნელობებთან ერთად, მიზანი არის t–ს მნიშვნელობის პროგნოზირება $\mathbf{x}$–ის ახალი მნიშვნელობისთვის. უმარტივესი მიდგომით, ეს შეიძლება გაკეთდეს $y(\mathbf{x})$ შესაბამისი ფუნქციის უშუალოდ აგებით, რომლის მნიშვნელობები $\mathbf{x}$–ის ახალი მნიშვნელობებისთვის წარმოადგენს პროგნოზებს t–ის შესაბამისი მნიშვნელობებისთვის. უფრო ზოგადად, ალბათური პერსპექტივიდან, ჩვენი მიზანია პროგნოზირებადი განაწილების $p(t\,|\,\mathbf{x})$

მოდელირება, რადგან ეს გამოხატავს ჩვენს გაურკვევლობას t –ს მნიშვნელობის შესახებ $\mathbf{x}$ –ის ყოველი მნიშვნელობისათვის. ამ პირობითი განაწილებიდან შეგვიძლია გავაკეთოთ t –ს პროგნოზირებები $\mathbf{x}$ –ის ნებისმიერი ახალი მნიშვნელობისთვის, ისე, რომ შევამციროთ სათანადოდ შერჩეული დანაკარგების ფუნქციის მოსალოდნელი მნიშვნელობა. რეალურ მნიშვნელობებიანი ცვლადებისათვის დანაკარგების ფუნქციის ზოგადი არჩევანია დანაკარგების კვადრატი, რომლისთვისაც ოპტიმალური ამოხსნა მოცემულია t –ს პირობითი მოლოდინით.

წრფივ მოდელებს გააჩნიათ მნიშვნელოვანი შეზღუდვები სახეთა გამოცნობის პრაქტიკულ გამოყენებებში, განსაკუთრებით მაღალ განზომილებიანი საწყისი მონაცემების (დამოუკიდებელი მონაცემების) სივრცის პრობლემების გამო, მაგრამ მათ აქვთ კარგი ანალიტიკური თვისებები და ქმნიან საფუძველს უფრო დახვეწილი მოდელების შესასწავლად.

რეგრესიის უმარტივესი წრფივი მოდელია ის, რაც გულისხმობს შეყვანის ცვლადების წრფივ კომბინაციას

$$y(\mathbf{x},\mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D , \qquad (9.1)$$

სადაც $\mathbf{x} = (x_1,\ldots,x_D)^T$. ეს ხშირად, უბრალოდ, ცნობილია *წრფივი რეგრესიის* სახელით. ამ მოდელის მთავარი თვისებაა, რომ ის არის $w_0,\ldots,w_D$, პარამეტრების წრფივი ფუნქცია. ამასთან, ეს არის x_i შეყვანის ცვლადების წრფივი ფუნქცია და ეს მნიშვნელოვან შეზღუდვებს ადებს მოდელს. ამიტომ მოდელების კლასს ვაფართოვებთ შემდეგი ფორმის შეყვანის ცვლადების ფიქსირებული არაწრფივი ფუნქციების წრფივი კომბინაციების განხილვითით

$$y(\mathbf{x},\mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) , \qquad (9.2)$$

სადაც $\phi_j(\mathbf{x})$ ცნობილია როგორც *ბაზისური ფუნქციები*. აღვნიშნოთ j ინდექსის მაქსიმალური მნიშვნელობა $M-1$ -ით. მაშინ, ამ მოდელის პარამეტრების მთლიანი რაოდენობა იქნება M .

პარამეტრი w_0 მონაცემთა განლაგების ფიქსირების საშუალებას იძლევა და ზოგჯერ მას უწოდებენ *კომპენსაციის* პარამეტრს. ხშირად, მოსახერხებელია დამატებითი ცრუ "ბაზისური ფუნქციის" განსაზღვრა $\phi_0(\mathbf{x}) = 1$ ისე, რომ

$$y(\mathbf{x},\mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) , \qquad (9.3)$$

სადაც $\mathbf{w} = (w_0,\ldots,w_{M-1})^T$ და $\phi = (\phi_0,\ldots,\phi_{M-1})^T$. სახეთა გამოცნობის მრავალ პრაქტიკულ გამოყენებებში, მონაცემთა ორიგინალი (საწყისი) ცვლადებისათვის ვიყენებთ ფიქსირებული წინასწარი დამუშავების ან თავისებურების გამოყოფის გარკვეულ ფორმას. თუ ორიგინალი

ცვლადები შეიცავენ $\mathbf{x}$ ვექტორს, მაშინ თავისებურებები შეიძლება გამოისახოს ბაზისური ფუნქციების $\{\phi_j(\mathbf{x})\}$ საშუალებით.

არაწრფივი ბაზისური ფუნქციების გამოყენებით საშუალება გვაქვს $y(\mathbf{x},\mathbf{w})$ ფუნქცია წარმოვადგინოთ, როგორც $\mathbf{x}$ შეყვანის ვექტორის არაწრფივი ფუნქცია. (9.2) ფორმის მქონე ფუნქციებს წრფივ მოდელებს უწოდებენ, თუმცა ეს ფუნქცია წრფივია $\mathbf{w}$–ს მიმართ. პარამეტრების სწორედ ეს წრფივობა მნიშვნელოვნად ამარტივებს ამ კლასის მოდელების ანალიზს. ამასთან, ეს ასევე იწვევს მნიშვნელოვან შეზღუდვებს.

პოლინომური რეგრესიის მაგალითი წარმოადგენს ამ მოდელის კერძო მაგალითს, რომელშიც არის x შეყვანის ერთი ცვლადი, და ბაზისური ფუნქციები არიან x–ის ხარისხები, ასე რომ $\phi_j(x) = x^j$. პოლინომიალური ბაზისური ფუნქციების ერთი შეზღუდვა ის არის, რომ ისინი არიან შეყვანის ცვლადის გლობალური ფუნქციები, ისე, რომ შეყვანის სივრცის ერთ არეში ცვლილებები გავლენას ახდენს ყველა სხვა არეებზე. ეს შეიძლება გადაწყდეს შეყვანის სივრცის არეებად დაყოფით და ყოველ არეში სხვადასხვა პოლინომის ჩასმით, რასაც მივყავართ *სპლაინ ფუნქციებთან.*

არსებობს ბაზისური ფუნქციების არჩევის მრავალი სხვა შესაძლებლობა, მაგალითად

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}, \qquad (9.4)$$

სადაც μ_j არეგულირებს ბაზისური ფუნქციების მდებარეობებს შეყვანის სივრცეში, ხოლო პარამეტრი s მართავს მათ სივრცულ მასშტაბს. ეს ჩვეულებრივ მოიხსენიება როგორც "გაუსის" ბაზისური ფუნქციები, თუმცა, უნდა აღინიშნოს, რომ არ მოითხოვება, რომ მათ ჰქონდეთ ალბათური ინტერპრეტაცია და, კერძოდ, ნორმალიზაციის კოეფიციენტი არ არის მნიშვნელოვანი, რადგან ეს ბაზისური ფუნქციები გამრავლდებიან w_j ადაპტაციურ პარამეტრებზე.

კიდევ ერთი შესაძლებლობა არის შემდეგი ფორმის სიგმოიდური ბაზისური ფუნქცია

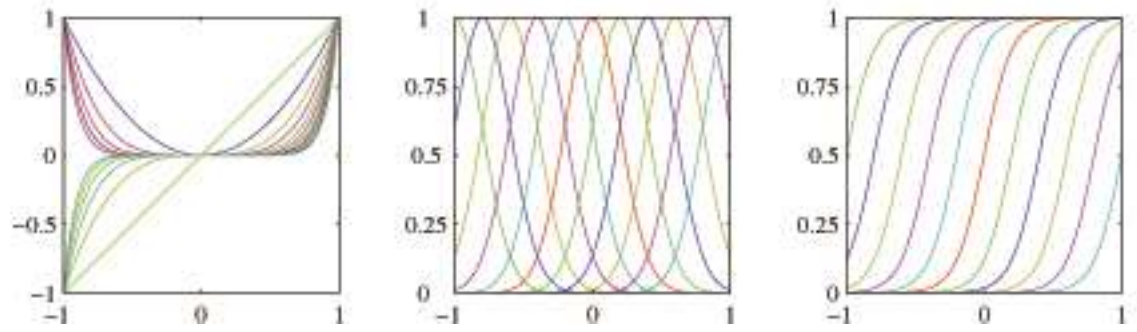
$$\phi_j(x) = \sigma\left(\frac{x-\mu_j}{s}\right), \qquad (9.5)$$

სადაც $\sigma(a)$ არის ლოგისტიკური სიგმოიდური ფუნქცია, განსაზღვრული ასე

$$\sigma(a) = \frac{1}{1+\exp(-a)}. \qquad (9.6)$$

ანალოგიურად, შეგვიძლია გამოვიყენოთ „tanh“ ფუნქცია, რადგან ეს დაკავშირებულია ლოგისტიკურ სიგმოიდთან დამოკიდებულებით $\tanh(a) = 2\sigma(2a) - 1$ და, ამიტომ,

ლოგისტიკური სიგმოიდური ფუნქციების ზოგადი წრფივი კომბინაცია ექვივალენტურია „tanh“ ფუნქციების ზოგადი წრფივი კომბინაციის. ბაზისური ფუნქციის ეს სხვადასხვა არჩევანი ასახულია ნახაზ 9.1–ზე.



ნახ. 9.1. ბაზისური ფუნქციების მაგალითები, მარცხნივ ნაჩვენებია პოლინომები, (9.4) ფორმის გაუსის – ცენტრში და (9.5) ფორმის სიგმოიდური – მარჯვნივ.

ბაზისური ფუნქციის კიდევ ერთი შესაძლო არჩევანი არის ფურიეს ბაზისი, რომელსაც მივყავართ სინუსოიდური ფუნქციებით გაშლასთან. ყოველი ბაზისური ფუნქცია წარმოადგენს სპეციფიკურ სიხშირეს და აქვს უსასრულო სივრცული დიაპაზონი. ამის საპირისპიროდ, ბაზისური ფუნქციები, რომლებიც ლოკალიზებულია შეყვანის სივრცის სასრულ არეებში, აუცილებლად მოიცავს სხვადასხვა სივრცული სიხშირის სპექტრს. სიგნალის დამუშავების მრავალ გამოყენებებში საინტერესოა ბაზისური ფუნქციების განხილვა, რომლებიც ლოკალიზებულია როგორც სივრცეში, ასევე სიხშირეში, რასაც მივყავართ ფუნქციების კლასთან, რომლებიც ცნობილია როგორც *ტალღური.* ისინი ასევე განსაზღვრულია ისე, რომ იყვნენ ორთოგონალური, რაც ამარტივებს მათ გამოყენებას. ტალღურები ყველაზე მეტად გამოიყენება, როდესაც შეყვანის მნიშვნელობები მოცემულია რეგულარულ ბადეზე, მაგალითად, დროის მიმდევრობითი წერტილები დროით მწკრივში ან პიქსელები გამოსახულებაში.

ამასთან, ამ თავში განხილვის უმეტესი ნაწილი დამოუკიდებელია ბაზისური ფუნქციების სიმრავლისგან. ამიტომ, განხილვის უმეტეს ნაწილში არ დავაკონკრეტებთ ბაზისური ფუნქციებების ფორმას, გარდა რიცხვითი ილუსტრაციის მიზნებისა. მართლაც, განხილვის უმეტესი ნაწილი ერთნაირად იქნება გამოყენებადი სიტუაციისთვის, რომელშიც ბაზისური ფუნქციების ვექტორი $\phi(\mathbf{x})$ არის უბრალო იდენტურობა $\phi(\mathbf{x})=\mathbf{x}$. გარდა ამისა, იმისათვის, რომ აღნიშვნები იყოს მარტივი, ყურადღებას გავამახვილებთ ერთი სამიზნე ცვლადის t–ს შემთხვევაზე. ამასთან, მოკლედ განვიხილავთ იმ მოდიფიკაციებს, რომლებიც საჭიროა მრავალგანზომილებიან სამიზნე ცვლადთან გასამკლავებლად.

თეორიულ საკითხებზე მუშაობის შემდეგ განხილული იქნება პრაქტიკული მაგალითები სემინარის თემის მიხედვით და მათი ამოხსნის მეთოდები.

სემინარული სამუშაო # 12

სემინარის თემა: კლასიფიკაციის წრფივი მოდელები.

სემინარზე განხილული იქნება სტატისტიკური წრფივი მოდელების კლასები კლასიფიკაციის პრობლემების გადასაჭრელად. ზოგადად, კლასიფიკაციის მიზანი არის შეყვანის $\mathbf{x}$ ვექტორის მიკუთვნება რომელიმე K დისკრეტული C_k კლასისათვის, სადაც $k=1,\ldots,K$. ყველაზე გავრცელებულ სცენარში კლასები არიან განცალკევებული, ისე რომ თითოეული შეყვანის ვექტორი მიეკუთვნება მხოლოდ ერთ კლასს. ამრიგად, შეყვანის სივრცე იყოფა *გადაწყვეტილების არეებად*, რომელთა საზღვრებს *გადაწყვეტილების საზღვრებს* ან *გადაწყვეტილების ზედაპირებს* უწოდებენ. სემინარზე განვიხილავთ კლასიფიკაციის წრფივ მოდელებს, რომლის ქვეშაც იგულისხმება, რომ გადაწყვეტილების ზედაპირები წარმოადგენენ $\mathbf{x}$ შეყვანის ვექტორის წრფივ ფუნქციებს და, შესაბამისად, განსაზღვრული არიან $(D-1)$-განზომილებიანი ჰიპერსიბრტყეებით D- განზომილებიან შეყვანის სივრცეში. ამ მიზნით განხილული იქნა შემდეგის საკითხები: დისკრიმინანტის ფუნქციები, ორი კლასის შემთხვევა, უმცირეს კვადრატი კლასიფიკაციისათვის, ფიშერის წრფივი დისკრიმინანტი, უმცირეს კვადრატებს შორის კავშირი და ფიშერის დისკრიმინანტი მრავლობითი კლასებისათვის. სემინარზე განხილული საკითხები დაკავშირებულია კურსის სახელმძღვანელოს „მანქანური სწავლება, სახელმძღვანელო (ლექციების კურსი) უნივერსიტეტის სტუდენტებისათვის“, მომზადებული ქ.ი. ყაჭიაშვილის მიერ წიგნის „M. Bishop, Pattern Recognition and Machine Learning. Springer Verlag“ გარკვეული პუნქტების თარგმანი, თავი 12-ში მოცემულ თეორიულ მასალასთან.

რეგრესიის პრობლემებისათვის სამიზნე ცვლადი $\mathbf{t}$ იყო უბრალოდ რეალური რიცხვების ვექტორი, რომელთა მნიშვნელობების პროგნოზირებაც გვსურდა. კლასიფიკაციის შემთხვევაში, სამიზნე მნიშვნელობების გამოყენების სხვადასხვა გზა არსებობს კლასის ჭდის წარმოსადგენად. ალბათური მოდელებისთვის ყველაზე მოსახერხებელია, ორკლასიანი პრობლემების შემთხვევაში, ორობითი წარმოდგენა, რომელშიც არის ერთი სამიზნე ცვლადი $t\in\{0,1\}$ ისეთი, რომ $t=1$ წარმოადგენს C_1 კლასს და $t=0$ წარმოადგენს C_2 კლასს. შეგვიძლია ინტერპრეტირება გავუკეთოთ t-ს მნიშვნელობას როგორც იმის ალბათობას, რომ კლასი არის C_1, ალბათობის მნიშვნელობებით რომლებიც მხოლოდ უკიდურეს 0 და 1 მნიშვნელობებს ღებულობან. $K>2$ კლასებისათვის, მოსახერხებელია გამოვიყენოს K–დან ერთის კოდირების სქემა, რომელშიც t არის K სიგრძის ვექტორი ისეთი, რომ თუ კლასი არის C_j, მაშინ $\mathbf{t}$–ს ყველა t_k ელემენტი ნულის ტოლია, გარდა t_j ელემენტისა, რომელიც იღებს მნიშვნელობას 1. მაგალითად, თუ გვაქვს $K=5$ კლასები, მაშინ მე-2 კლასის სახე მოიცემა სამიზნე ვექტორით

$$\mathbf{t}=(0,1,0,0,0)^T. \qquad (12.1)$$

როგორც ვთქვით, t_k-ს მნიშვნელობა შეგვიძლია განვიხილოთ როგორც ალბათობა იმისა, რომ კლასი არის C_k. არა ალბათური მოდელებისათვის, სამიზნე ცვლადის წარმოდგენის ალტერნატიული არჩევანი ზოგჯერ მოხერხებული აღმოჩნდება.

მე–4 თავში განვსაზღვრეთ კლასიფიკაციის პრობლემის სამი განსხვავებული მიდგომა. უმარტივესი გულისხმობს დისკრიმინაციული ფუნქციის აგებას, რომელიც პირდაპირ ანიჭებს თითოეულ $\mathbf{x}$ ვექტორს კონკრეტულ კლასს. უფრო ძლიერი მიდგომა მოდელირებას უკეთებს ალბათობის პირობით განაწილებას $p(C_k \mid \mathbf{x})$ შეფასების ეტაპზე, შემდეგ კი იყენებს ამ განაწილებას ოპტიმალური გადაწყვეტილებების მისაღებად. შეფასების და გადაწყვეტილების განცალკევებით ვღებულობთ უამრავ სარგებელს, როგორც ეს განხილულია § 4.5–ში. პირობითი ალბათობების $p(C_k \mid \mathbf{x})$ განსაზღვრის ორი განსხვავებული მიდგომა არსებობს. ერთი ტექნიკაა მათი უშუალო მოდელირება, მაგალითად, მათი წარმოდგენა პარამეტრულ მოდელებად და შემდეგ პარამეტრების ოპტიმიზაცია სასწავლო სიმრავლის გამოყენებით. გარდა ამისა, შეგვიძლია გამოვიყენოთ გენერაციული მიდგომა, რომელშიც მოდელირებას ვუკეთებთ $p(\mathbf{x} \mid C_k)$–ით მოცემულ კლას-პირობით სიმკვრივეებს, კლასებისთვის $p(C_k)$ აპრიორულ ალბათობებთან ერთად, და შემდეგ კი გამოვითვლით საჭირო აპოსტერიორულ ალბათობებს ბაიესის თეორემის გამოყენებით

$$p(C_k \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_k)\, p(C_k)}{p(\mathbf{x})}. \qquad (12.2)$$

ამ თავში სამივე მიდგომის მაგალითებს განვიხილავთ.

მე–9 თავში განხილული წრფივი რეგრესიის მოდელებში მოდელის პროგნოზი $y(\mathbf{x}, \mathbf{w})$ მოცემული იყო $\mathbf{w}$ პარამეტრების წრფივი ფუნქციით. უმარტივეს შემთხვევაში მოდელი ასევე წრფივია შეყვანის ცვლადების მიმართ და ამიტომ იღებს ფორმას $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$, ისე რომ y არის ნამდვილი რიცხვი. კლასიფიკაციის პრობლემების შემთხვევაში გვსურს განვსაზღვროთ კლასის დისკრეტული ჭდეები, ანუ უფრო ზოგადად, აპოსტერიოტული ალბათობები, რომლებიც იმყოფებიან დიაპაზონში $(0,1)$. ამის მისაღწევად განვიხილავთ ამ მოდელის განზოგადებას, რომელშიც გარდავქმნით $\mathbf{w}$–ს წრფივ ფუნქციას არაწრფივი $f(\cdot)$ ფუნქციის გამოყენებით ისე, რომ

$$y(\mathbf{x}) = f(\mathbf{w}^T\mathbf{x} + w_0). \qquad (12.3)$$

მანქანური სწავლის ლიტერატურაში $f(\cdot)$ ცნობილია, როგორც *აქტივაციის ფუნქცია*, ხოლო მის ინვერსიას (შებრუნებულს) სტატისტიკურ ლიტერატურაში *კავშირის ფუნქცია* ეწოდება. გადაწყვეტილების ზედაპირები შეესაბამება $y(\mathbf{x}) =$constant–ს, ასე, რომ $\mathbf{w}^T\mathbf{x} + w_0 =$constant და, შესაბამისად, გადაწყვეტილების ზედაპირები არიან $\mathbf{x}$-ის წრფივი ფუნქციები, თუნდაც $f(\cdot)$ ფუნქცია არაწრფივი იყოს. ამ მიზეზით, (12.3)–ით აღწერილი მოდელების კლასს *განზოგადებული წრფივი მოდელები* ეწოდება. ამასთან, შევნიშნოთ, რომ რეგრესიისთვის გამოყენებული მოდელებისგან განსხვავებით, ისინი პარამეტრებით წრფივი აღარ არიან არაწრფივი $f(\cdot)$ ფუნქციის არსებობის გამო. ამას მივყევართ უფრო რთულ ანალიზურ და გამოთვლით თვისებებთან ვიდრე წრფივ რეგრესიის მოდელებს აქვთ. ამის მიუხედავად, ეს მოდელები მაინც შედარებით მარტივია უფრო ზოგად არაწრფივ მოდელებთან შედარებით.

ამ თავში განხილული ალგორითმები თანაბრად იქნებიან გამოყენებადი, თუ პირველად განვახორციელებთ შეყვანის ცვლადების ფიქსირებულ არაწრფივ გარდაქმნას $\phi(\mathbf{x})$ საბაზისო ფუნქციების ვექტორის გამოყენებით, როგორც ეს რეგრესიის მოდელებისთვის გავაკეთეთ მე-9 თავში. დავიწყებთ კლასიფიკაციის განხილვით უშუალოდ ორიგინალურ $\mathbf{x}$ შეყვანის სივრცეში.

დისკრიმინატორი არის ფუნქცია, რომელიც იღებს $\mathbf{x}$ შეყვანის ვექტორს და მიაწერს მას ერთ–ერთს K კლასიდან, რომელიც აღინიშნება C_k–თი. ამ თავში შემოვიფარგლებით *წრფივი დისკრიმინანტების* განხილვით, კერძოდ, ისეთებით, რომლებისთვისაც გადაწყვეტილების ზედაპირები არიან ჰიპერსიბრტყეები. განხილვის გამარტივების მიზნით, პირველ რიგში განვიხილავთ ორი კლასის შემთხვევას, შემდეგ კი გამოვიკვლევთ განზოგადოებას $K>2$ კლასებისათვის.

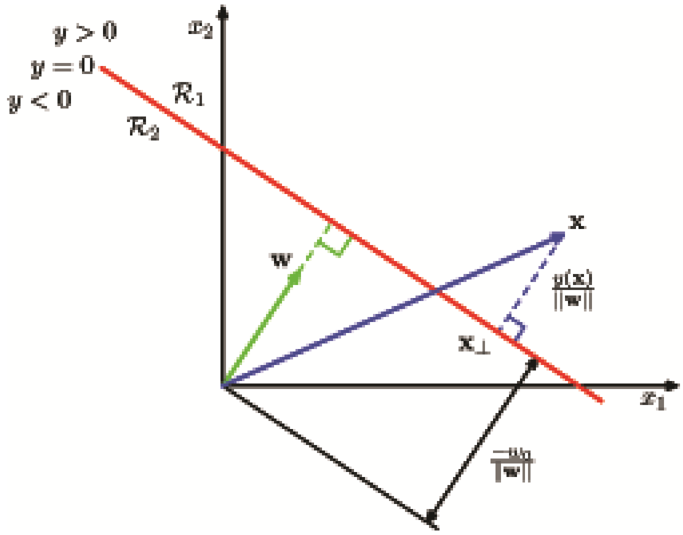
წრფივი დისკრიმინანტული ფუნქციის უმარტივესი წარმოდგენა მიიღება შეყვანის ვექტორის წრფივი ფუნქციის აღებით ისე, რომ

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0, \qquad (12.4)$$

სადაც $\mathbf{w}$–ს *წონის ვექტორი* ეწოდება და w_0 არის კომპენსაცია. კომპენსაციის უარყოფით მნიშვნელობას ზოგჯერ *ზღურბლს* უწოდებენ. შეყვანის ვექტორი $\mathbf{x}$ მიეკუთვნება C_1 კლასს, თუ $y(\mathbf{x}) \geq 0$ და C_2 კლასს სხვა შემთხვევაში. შესაბამისი გადაწყვეტილების საზღვარი განისაზღვრება $y(\mathbf{x}) = 0$ დამოკიდებულებით, რომელიც შეესაბამება $(D-1)$-განზომილებიანს ჰიპერსიბრტყეს D-განზომილებიან შეყვანის სივრცეში. განვიხილოთ ორი წერტილი $\mathbf{x}_A$ და $\mathbf{x}_B$, რომლებიც მდებარეობენ გადაწყვეტილების ზედაპირზე. იმის გამო, რომ $y(\mathbf{x}_A) = y(\mathbf{x}_B) = 0$, გვაქვს $\mathbf{w}^T(\mathbf{x}_A - \mathbf{x}_B) = 0$ და, შესაბამისად, ვექტორი $\mathbf{w}$ ორთოგონალურია ყველა ვექტორის მიმართ, რომლებიც მდებარეობენ გადაწყვეტილების ზედაპირზე და ამდენად $\mathbf{w}$ განსაზღვრავს გადაწყვეტილების ზედაპირის ორიენტაციას. ანალოგიურად, თუ $\mathbf{x}$ არის წერტილი გადაწყვეტილების ზედაპირზე, მაშინ $y(\mathbf{x}) = 0$ და, ამდენად, ნორმალური მანძილი სათავიდან (კოორდინატთა სათავიდან) გადაწყვეტილების ზედაპირამდე მოიცემა ასე

$$\frac{\mathbf{w}^T\mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|}. \qquad (12.5)$$

ამიტომ ვხედავთ, რომ კომპენსაციის პარამეტრი w_0 განსაზღვრავს გადაწყვეტილების ზედაპირის მდებარეობას. ეს თვისებები ილუსტრირებულია ნახაზ 12.1–ზე $D=2$ შემთხვევისთვის.



ნახ. 12.1. წრფივი დისკრიმინანტული ფუნქციის გეომეტრიის ილუსტრაცია ორ–განზომილებაიან შემთხვევაში. გადაწყვეტილების ზედაპირი, რომელიც წითლად არის ნაჩვენები, $\mathbf{w}$-ს პერპენდიკულარულია და მისი გადაადგილება კოორდინატთა სათავიდან კონტროლდება w_0 წანაცვლების პარამეტრით. ასევე, ზოგადი $\mathbf{x}$ წერტილის აღნიშნული ორთოგონალური მანძილი გადაწყვეტილების ზედაპირიდან მოცემულია $y(\mathbf{x})/\|\mathbf{w}\|$ -ით.

გარდა ამისა, აღვნიშნავთ, რომ $y(\mathbf{x})$–ის მნიშვნელობა იძლევა $\mathbf{x}$ წერტილის r პერპენდიკულარული მანძილის აღნიშნულ ზომას გადაწყვეტილების ზედაპირიდან. ამის სანახავად განვიხილოთ ნებისმიერი $\mathbf{x}$ წერტილი და $\mathbf{x}_\perp$ იყოს მისი ორთოგონალური პროექცია გადაწყვეტილების ზედაპირზე ისე, რომ

$$\mathbf{x} = \mathbf{x}_\perp + r\frac{\mathbf{w}}{\|\mathbf{w}\|}. \qquad (12.6)$$

ამ შედეგის ორივე მხარის $\mathbf{w}^T$-ზე გამრავლებით და w_0–ის დამატებით, აგრეთვე $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$ და $y(\mathbf{x}_\perp) = \mathbf{w}^T\mathbf{x}_\perp + w_0 = 0$ გამოყენებით, გვაქვს

$$r = \frac{y(\mathbf{x})}{\|\mathbf{w}\|}. \qquad (12.7)$$

ეს შედეგი ილუსტრირებულია ნახაზ 12.1–ზე.

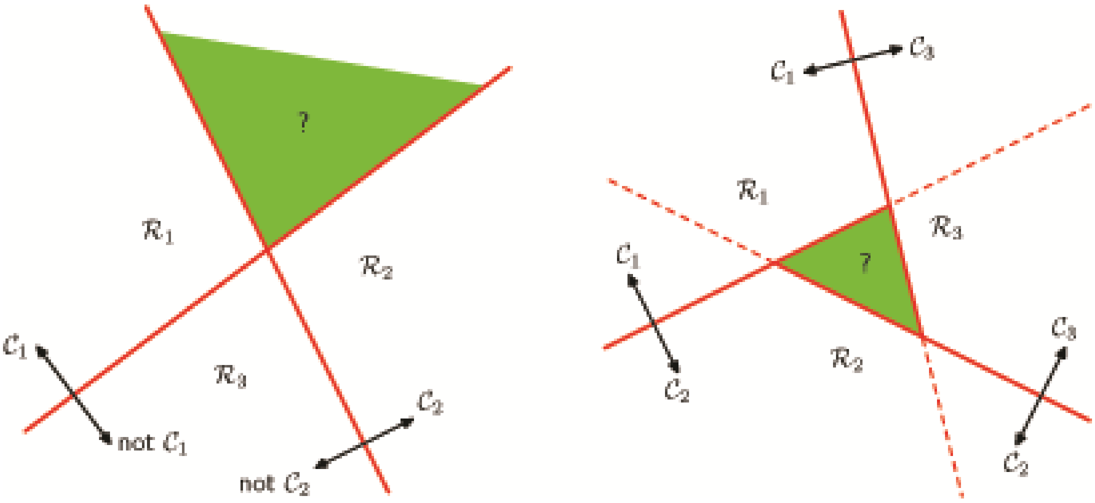
მე–9 თავში განხილული წრფივი რეგრესიული მოდელების მსგავსად, ზოგჯერ უფრო მოსახერხებელია გამოვიყენოთ უფრო კომპაქტური აღნიშვნა, რომელშიც შემოვიღებთ დამატებით ცრუ "შეყვანის" მნიშვნელობას $x_0 = 1$ და შემდეგ განვსაზღვრავთ $\widetilde{\mathbf{w}} = (w_0, \mathbf{w})$ და $\widetilde{\mathbf{x}} = (x_0, \mathbf{x})$ ისე, რომ

$$y(\mathbf{x}) = \widetilde{\mathbf{w}}^T\widetilde{\mathbf{x}}. \qquad (12.8)$$

ამ შემთხვევაში, გადაწყვეტილების ზედაპირები არიან D–განზომილებიანი ჰიპერსიბრტყეები, რომლებიც გადიან $D+1$ განზომილებიანი გაფართოებული შეყვანის სივრცის საწყის წერტილზე.

ახლა განიხილეთ წრფივი დისკრიმინანტების განზოგადება $K>2$ კლასებზე. შეიძლება გაგვიჩნდეს ცდუნება ავაგოთ K კლასის დისკრიმინატორი, ორი–ორი კლასების დისკრიმინანტორი ფუნქციების გაერთიანებით. თუმცა, ეს იწვევს სერიოზულ სირთულეებს, როგორც იქნება ნაჩვენები ქვემოთ.

განვიხილოთ $K-1$ რაოდენობის კლასიფიკატორების გამოყენება, რომელთაგან თითოეულ შემთხვევაში წყდება ორი კლასის პრობლემა, წერტილების მიკუთვნების ან არ მიკუთვნების კონკრეტული C_k კლასისათვის. ეს ცნობილია როგორც *ერთი დანარჩენების წინააღმდეგ* კლასიფიკატორი. ნახაზ 12.2-ზე მოცემული მარცხენა მაგალითი გვიჩვენებს შემთხვევას, რომელიც მოიცავს სამ კლასს, სადაც ამ მიდგომას მივყავართ შეყვანის სივრცის არეებთან, რომლებიც ორაზროვნად არიან კლასიფიცირებული.



ნახ. 12.2. K კლასის დისკრიმინანტის კონსტრუქციის მცდელობა ორი კლასის დისკრიმინანთა სიმრავლიდან მივყავართ ორაზროვან არეებთან, რომლებიც ნაჩვენებია მწვანე ფერში. მარცხნივ არის მაგალითი, რომელიც მოიცავს ორი დისკრიმინანტის გამოყენებას, რომლებიც შექმნილია C_k კლასის წერტილების განსასხვავებლად წერტილებისაგან, რომლებიც არ მიეკუთვნებიან C_k კლასს. მარჯვნივ მოცემულია მაგალითი, რომელიც მოიცავს სამ დისკრიმინანტულ ფუნქციას, რომელთაგან თითოეული გამოიყენება C_k და C_j კლასების განცალკევებისათვის.

ალტერნატივა არის $K(K-1)/2$ ორობითი დისკრიმინანტული ფუნქციების შემოტანა, კლასების ყველა შესაძლო წყვილებისათვის თითო დისკრიმინანტული ფუნქცია. ეს ცნობილია როგორც *ერთი-ერთის წინააღმდეგ* კლასიფიკატორი. შემდეგ თითოეული წერტილი კლასიფიცირდება დისკრიმინანტური ფუნქციების „ხმების“ უმრავლესობის მიხედვით. ამასთან, აქაც ბუნდოვანი არეების პრობლემას წარმოიშობა, როგორც ეს ნაჩვენებია ნახაზ 12.2-ის მარჯვენა დიაგრამაზე.

ამ სირთულეების თავიდან აცილება შეგვიძლია K-კლასის ერთი დისკრიმინატორის განხილვით, რომელიც შეიცავს შემდეგი სახის K წრფივ ფუნქციებს

$$y_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0}, \tag{12.9}$$

და შემდეგ $\mathbf{x}$ წერტილის მინიჭება C_k კლასს, თუ $y_k(\mathbf{x}) > y_j(\mathbf{x})$ ყველა $j \neq k$. გადაწყვეტილების საზღვარი C_k კლასსა და C_j კლასს შორის მოცემულია ტოლობით $y_k(\mathbf{x}) = y_j(\mathbf{x})$ და შესაბამისად შეესაბამება $(D-1)$-განზომილებიან ჰიპერსიბრტყეს, რომელიც განსაზღვრულია განტოლებით

$$(\mathbf{w}_k - \mathbf{w}_j)^T \mathbf{x} + (w_{k0} - w_{j0}) = 0. \tag{12.10}$$

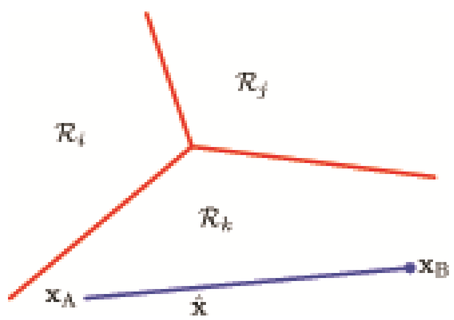
ამას აქვს იგივე ფორმა, როგორიც გადაწყვეტილების საზღვარს ორი კლასიანი შემთხვევისთვის და, შესაბამისად, ანალოგიური გეომეტრიული თვისებები გამოიყენება.

ასეთი დისკრიმინანტის გადაწყვეტილების არეები ყოველთვის მარტივად არიან ერთმანეთთან დაკავშირებული და ამოზნექილი. ამის სანახავად, განვიხილოთ ორი წერტილი $\mathbf{x}_A$ და $\mathbf{x}_B$, რომლებიც მდებარეობენ გადაწყვეტილების R_k არეში, როგორც ეს ნაჩვენებია ნახაზ 12.3–ზე. ნებისმიერი წერტილი $\hat{\mathbf{x}}$, რომელიც მდებარეობს $\mathbf{x}_A$ და $\mathbf{x}_B$ დამაკავშირებელ წრფეზე, შეიძლება გამოისახოს შემდეგი ფორმით

$$\hat{\mathbf{x}} = \lambda \mathbf{x}_A + (1-\lambda)\mathbf{x}_B. \tag{12.11}$$

სადაც $0 \le \lambda \le 1$. დისკრიმინაციული ფუნქციების წრფივობიდან გამომდინარეობს, რომ

$$y_k(\hat{\mathbf{x}}) = \lambda y_k(\mathbf{x}_A) + (1-\lambda) y_k(\mathbf{x}_B). \tag{12.12}$$



ნახ. 12.3. გადაწყვეტილების არეების ილუსტრაცია მრავალკლასიანი წრფივი დისკრიმინანტისთვის, წითლად ნაჩვენები გადაწყვეტილების საზღვრებით. თუ ორი წერტილი $\mathbf{x}_A$ და $\mathbf{x}_B$ ერთი და იგივე R_k გადაწყვეტილების არეში მდებარეობს, მაშინ ნებისმიერი $\hat{\mathbf{x}}$ წერტილი, რომელიც მდებარეობს ამ ორი წერტილის დამაკავშირებელ წრფეზე, ასევე უნდა განლაგდეს R_k – ში, შესაბამისად, გადაწყვეტილების არე უნდა იყოს მარტივად დაკავშირებული (შეერთებული) და ამოზნექილი.

რადგან ორივე $\mathbf{x}_A$ და $\mathbf{x}_B$ მდებარეობენ R_k -ს შიგნით, ამიტომ $y_k(\mathbf{x}_A) > y_j(\mathbf{x}_A)$ და $y_k(\mathbf{x}_B) > y_j(\mathbf{x}_B)$, ყველა $j \neq k$ და, შესაბამისად, $y_k(\hat{\mathbf{x}}) > y_j(\hat{\mathbf{x}})$ და ამდენად $\hat{\mathbf{x}}$ ასევე მდებარეობს R_k –ს შიგნით. ამრიგად, R_k არის მარტივად დაკავშირებული (შეერთებული) და ამოზნექილი.

შევნიშნოთ, რომ ორი კლასისთვის შეგვიძლია გამოვიყენოთ აქ განხილული ფორმალიზმი $y_1(\mathbf{x})$ და $y_2(\mathbf{x})$ ორი დისკრიმინაციული ფუნქციის საფუძველზე, ან გამოვიყენოთ უფრო მარტივი, მაგრამ ექვივალენტური ფორმულირება ერთი დისკრიმინანტული $y(\mathbf{x})$ ფუნქციის საფუძველზე.

აგრეთვე განვიხილავთ წრფივი დისკრიმინანტული ფუნქციების პარამეტრების შესწავლის სამ მიდგომას, დაფუძნებულს უმცირეს კვადრატებზე, ფიშერის წრფივ დისკრიმინანტზე და პერცეპტრონის ალგორითმზე.

მოდელებში, რომლებიც პარამეტრების წრფივი ფუნქციებია, შეცდომების კვადრატების ჯამის ფუნქციის მინიმიზაციას მივყავართ მარტივი ანალიტიკური ფორმის ამოხსნასთან პარამეტრების მნიშვნელობებისთვის. ამიტომ საინტერესოა იმის ნახვა, შეგვიძლია თუ არა იგივე ფორმალიზმის გამოყენება კლასიფიკაციის პრობლემებისათვის. განვიხილოთ ზოგადი კლასიფიკაციის პრობლემა K კლასებისათვის, K –დან 1-ის ორობითი კოდირების სქემის გამოყენებით სამიზნე $\mathbf{t}$ ვექტორისათვის. ასეთ კონტექსტში უმცირეს კვადრატთა გამოყენების ერთ-ერთი გამართლება არის ის, რომ ის აპროქსიმაციას უკეთებს სამიზნე მნიშვნელობების პირობით მათემატიკურ მოლოდინს $E[\mathbf{t} \mid \mathbf{x}]$ მოცემული შეყვანის ვექტორისათვის. ორობითი კოდირების სქემისთვის ეს პირობითი მოლოდინი მოცემულია კლასების აპოსტერიორული ალბათობების ვექტორით. სამწუხაროდ, ეს ალბათობები, როგორც წესი, საკმაოდ ცუდად არიან აპროქსიმირებული, მართლაც, მიახლოებებს შეიძლება ჰქონდეთ მნიშვნელობები $(0,1)$ დიაპაზონის ფარგლებს გარეთ წრფივი მოდელის შეზღუდული მოქნილობის გამო.

თითოეული C_k კლასი აღწერილია საკუთარი წრფივი მოდელით ისე, რომ

$$y_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0}, \qquad (12.13)$$

სადაც $k = 1, \ldots, K$. შესაძლებელია მათი მოხერხებულად დაჯგუფება ვექტორული აღნიშვნის გამოყენებით ისე, რომ

$$y(\mathbf{x}) = \widetilde{\mathbf{W}}^T \widetilde{\mathbf{x}}, \qquad (12.14)$$

სადაც $\widetilde{\mathbf{W}}$ არის მატრიცა, რომლის k –ური სვეტი შეიცავს $D+1$ განზომილებიან ვექტორს $\widetilde{\mathbf{w}}_k = \left(w_{k0}, \mathbf{w}_k^T\right)^T$ და $\widetilde{\mathbf{x}}$ არის შესაბამისი გაზრდილი შეყვანის ვექტორი $\left(1, \mathbf{x}^T\right)^T$ ხელოვნური შეყვანით $x_0 = 1$. ახალი $\mathbf{x}$ შეყვანა მიეკუთვნება კლასს, რომლისთვისაც გამომავალი (შედეგი) $y_k = \widetilde{\mathbf{w}}_k^T \widetilde{\mathbf{x}}$ არის ყველაზე დიდი.

ეხლა განვსაზღვრავთ პარამეტრების მატრიცას $\widetilde{\mathbf{W}}$ შეცდომების კვადრატების ჯამის ფუნქციის მინიმიზაციით. განვიხილოთ სასწავლო მონაცემთა სიმრავლე $\{\mathbf{x}_n, \mathbf{t}_n\}$, სადაც $n = 1, \ldots, N$, და

განვსაზღვროთ $\mathbf{T}$ მატრიცა, რომლის n–ური სტრიქონი არის $\mathbf{t}_n^T$ ვექტორი $\widetilde{\mathbf{X}}$ მატრიცასთან ერთად, რომლის n–ური სტრიქონი არის $\widetilde{\mathbf{x}}_n^T$. შეცდომების კვადრატების ჯამის ფუნქცია შეიძლება ჩაიწეროს შემდეგნაირად

$$E_D(\widetilde{\mathbf{W}}) = \frac{1}{2} Tr\left\{\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T}\right)^T \left(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T}\right)\right\}. \qquad (12.15)$$

$\widetilde{\mathbf{W}}$–თი წარმოებულის ნოლთან გატოლებით და გადალაგებით (მოწესრიგებით), მივიღებთ $\widetilde{\mathbf{W}}$–თი ამონახსნს შემდეგი სახით

$$\widetilde{\mathbf{W}} = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\mathbf{T} = \widetilde{\mathbf{X}}^{-}\mathbf{T}, \qquad (12.16)$$

სადაც $\widetilde{\mathbf{X}}^{-}$ არის $\widetilde{\mathbf{X}}$ მატრიცის ფსევდო-შებრუნებული. შემდეგ მივიღებთ დისკრიმინატულ ფუნქციას შემდეგი სახით

$$\mathbf{y}(\mathbf{x}) = \widetilde{\mathbf{W}}^T\widetilde{\mathbf{x}} = \mathbf{T}^T\left(\widetilde{\mathbf{X}}^{-}\right)^T\widetilde{\mathbf{x}}. \qquad (12.17)$$

უმცირეს კვადრატთა ამოხსნების საინტერესო თვისება მრავალი სამიზნე ცვლადით არის ის, რომ თუ ყველა სამიზნე ვექტორი სასწავლო სიმრავლეში აკმაყოფილებს გარკვეულ წრფივ შეზღუდვას

$$\mathbf{a}^T\mathbf{t}_n + b = 0, \qquad (12.18)$$

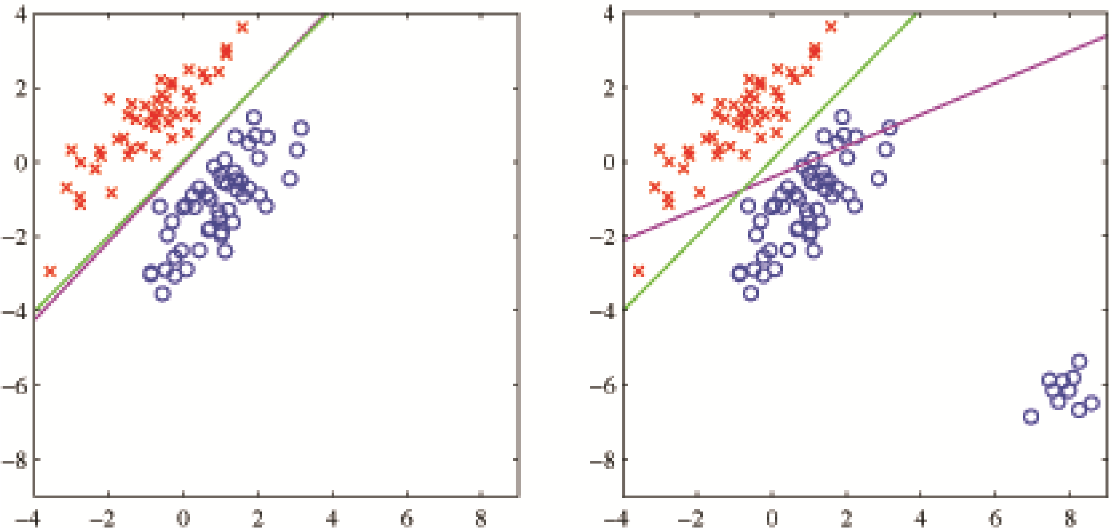
რაღაც $\mathbf{a}$ და b მუდმივზე, მაშინ მოდელის პროგნოზი $\mathbf{x}$–ის ნებისმიერი მნიშვნელობისათვის დააკმაყოფილებს იგივე შეზღუდვას ისე, რომ

$$\mathbf{a}^T\mathbf{y}(\mathbf{x}) + b = 0. \qquad (12.19)$$

ამრიგად, თუ K კლასებისთვის ვიყენებთ K–დან 1-ის კოდირების სქემას, მაშინ მოდელის მიერ გაკეთებულ პროგნოზებს ექნება თვისება, რომ $\mathbf{y}(\mathbf{x})$–ის ელემენტების ჯამი იკრიბება 1-კენ $\mathbf{x}$–ის ნებისმიერი მნიშვნელობისთვის. ამასთან, მხოლოდ ამ აჯამვის შეზღუდვა არ არის საკმარისი იმისთვის, რომ მოდელის შედეგები ინტერპრეტირებული იყოს როგორც ალბათობა, რადგან ისინი არ არიან შეზღუდული, რომ მიიღონ მნიშვნელობები მხოლოდ $(0,1)$ ინტერვალიდან.

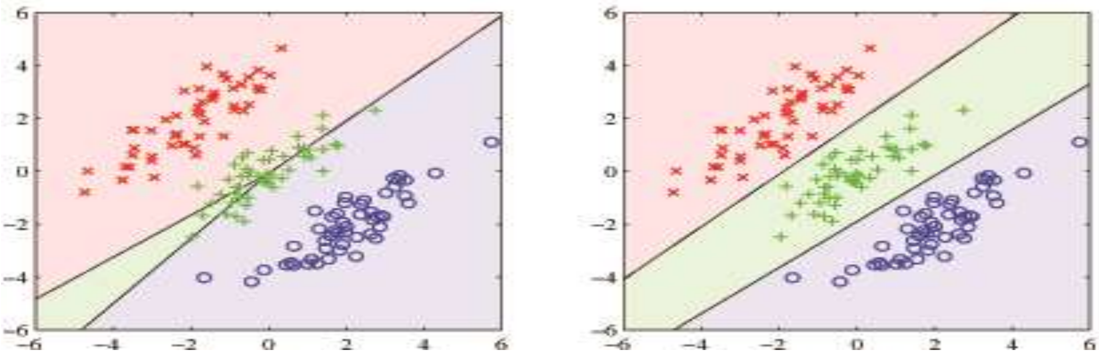
უმცირეს კვადრატთა მიდგომა იძლევა ზუსტი ანალიტიკური ფორმის ამოხსნას დისკრიმინანტული ფუნქციის პარამეტრებისთვის. ამასთან, როგორც დისკრიმინანტული ფუნქცია (სადაც მას ვიყენებთ უშუალოდ გადაწყვეტილების მისაღებად და გვერდს ვუვლით რაიმე ალბათურ ინტერპრეტაციას) მას გააჩნია გარკვეული სერიოზული პრობლემები. უკვე ვნახეთ, რომ უმცირეს კვადრატთა ამონახსნები არ არიან რობასტულები და ეს თანაბრად ეხება კლასიფიკაციისათვის მის გამოყენებასაც, როგორც ეს ნაჩვენებია ნახაზ 12.4–ზე. აქ ვხედავთ, რომ დამატებითი მონაცემების წერტილები მარჯვენა ფიგურაში მნიშვნელოვნად ცვლის გადაწყვეტილების საზღვრის ადგილმდებარეობას, მიუხედავად იმისა, რომ ეს წერტილი სწორად

კლასიფიცირდება მარცხენა ფიგურაში თავდაპირველი გადაწყვეტილების საზღვრის მიხედვით. შეცდომების კვადრატების ჯამის ფუნქცია აჯარიმებს პროგნოზებს, რომლებიც "ძალიან სწორია" იმით, რომ ისინი ლაგდებიან გადაწყვეტილების საზღვრის სწორე მხარის გრძელ გზაზე.



ნახ. 12.4. მარცხენა ნახაზი გვიჩვენებს მონაცემებს ორი კლასიდან, რომლებიც აღნიშნულია წითელი ჯვრებით და ლურჯი წრეებით, გადაწყვეტილების საზღვართან ერთად, რომელიც ნაპოვნია უმცირეს კვადრატთა მეთოდით (იისფერი მრუდი) და აგრეთვე ლოგისტიკური რეგრესიის მოდელით (მწვანე მრუდი), რომელსაც ამ კურსში არ განვიხილავთ. მარჯვენა ნახაზი გვიჩვენებს შესაბამის შედეგებს, როდესაც დიაგრამის მარცხენა ქვედა ნაწილში დამატებითი მონაცემთა წერტილებია დამატებული, რაც გვიჩვენებს, რომ უმცირეს კვადრატთა მეთოდი არის ძალიან მგრძნობიარე შედეგებთან დაკავშირებით, ლოგისტიკური რეგრესისგან განსხვავებით.

ამასთან, უმცირეს კვადრატთა მეთოდის პრობლემები შეიძლება იყოს უფრო მწვავე, ვიდრე უბრალოდ რობასტულობის არარსებობაა, როგორც ეს ნაჩვენებია ნახაზ 12.5–ზე. ამას გვიჩვენებს სინთეზური მონაცემების სიმრავლეები, რომლებიც მიღებულია სამი კლასიდან ორგანზომილებიან შეყვანის სივრცეში (x_1, x_2), რომელსაც აქვს თვისება, რომ წრფივ გადაწყვეტილების საზღვრებს შეუძლიათ მოახდინონ შესანიშნავი გამიჯვნა კლასებს შორის. მართლაც, ლოგისტიკური რეგრესიის ტექნიკა გვაძლევს დამაკმაყოფილებელ ამოხსნას, როგორც ეს ჩანს მარჯვენა სურათზე. ამასთან, უმცირეს კვადრატთა ამოხსნა იძლევა ცუდ შედეგს მწვანე კლასისთვის მიკუთვნებული შეყვანის სივრცის მცირე არის გამო.



ნახ. 12.5. სინთეზური მონაცემების სიმრავლის მაგალითი, რომელიც მოიცავს სამ კლასს, სასწავლო მონაცემების წერტილებით, რომლებიც აღინიშნულია წითლად (×), მწვანედ (+) და ლურჯად (○). ხაზები აღნიშნავს გადაწყვეტილების საზღვრებს, ხოლო ფონის ფერები აღნიშნავენ შესაბამისი კლასების გადაწყვეტილების არეებს. მარცხნივ არის უმცირეს კვადრატთა დისკრიმინანტის გამოყენების შედეგი. ვხედავთ, რომ მწვანე კლასისთვის გამოყოფილი შეყვანის სივრცის არე ძალიან მცირეა და ამიტომ ამ კლასის წერტილების უმეტესობა არასწორად არის კლასიფიცირებული. მარჯვნივ მოცემულია ლოგისტიკური რეგრესიების გამოყენების შედეგი (რომელსაც ამ კურსში არ განვიხილავთ), რომელიც გვიჩვენებს სასწავლო მონაცემების სწორ კლასიფიკაციას.

უმცირეს კვადრატთა ნაკლი არ უნდა გაგვიკვირდეს თუ გავიხსენებთ, რომ ის შეესაბამება მაქსიმალურ დასაჯერობას გაუსის პირობითი განაწილებისათვის, მაშინ როდესაც ორობითი სამიზნე ვექტორებს აშკარად აქვთ განაწილება, რომელიც შორს არის გაუსის განაწილებისაგან. უფრო შესაფერისი ალბათური მოდელების გამოყენებით მივიღებთ კლასიფიკაციის ტექნიკას ბევრად უკეთესი თვისებებით, ვიდრე უმცირეს კვადრატთა მეთოდია. ეხლა გავაგრძელოთ წრფივი კლასიფიკაციის მოდელებში პარამეტრების პოვნის ალტერნატიული, არა ალბათური, მეთოდების განხილვა.

თეორიულ საკითხებზე მუშაობის შემდეგ განხილული იქნება პრაქტიკული მაგალითები სემინარის თემის მიხედვით და მათი ამოხსნის მეთოდები.

რეკომენდებული ლიტერატურა:

1. „მანქანური სწავლება, სახელმძღვანელო (ლექციების კურსი) უნივერსიტეტის სტუდენტებისათვის“, მომზადებული ქ.ი. ყაჯიაშვილის მიერ წიგნის „M. Bishop, Pattern Recognition and Machine Learning. Springer Verlag“ გარკვეული პუნქტების თარგმანი.

2. Ch. M. Bishop. Pattern Recognition and Machine Learning. Springer Verlag, 2006.

3. J.M. Buhmann. Machine Learning, Swis Federal Institute of Technology, Zurich. 2015. CD 3211

4. Alpaydin, Ethem., “Introduction to machine learning”- 2nd ed. , The MIT Press 2009, ISBN 978-0-262-01243-0, CD 3238.

გადაეცა წარმოებას 30.03.2021. ხელმოწერილია დასაბეჭდად 6.04.2021. ოფსეტური ქაღალდის ზომა 60X84 1/16. პირობითი ნაბეჭდი თაბახი 2,5. ტირაჟი 50 ეგზ.

Verba volant, scripta manent

სტუ-ს „IT კონსალტინგის სამეცნიერო ცენტრი“, თბილისი, მ.კოსტავას 77

ISBN 978-9941-8-1753-3