# კლასტერული ანალიზი

**კლასტერული ანალიზის** ძირითად მიზანს **მონაცემთა სტრუქტურის გამოვლენა და ანალიზი წარმოადგენს**. **კლასტერული ანალიზის** ფარგლებში **მონაცემთა სტრუქტურის კვლევა კლასტერიზაციის –** დაჯგუფების საფუძველზე ხდება.

კლასტერიზაციის შედეგად გამოყოფილ ობიექტთა ჯგუფებს **კლასტერები**[1] ეწოდება. ცნება "**კლასტერის**" სინონიმებად იყენებენ "**ტაქსონს**", "**კლასს**", "**ჯგუფს**", "**სიმრავლეს**", "**ერთობლიობას**" და ა.შ.

კლასტერული ანალიზის მეშვეობით ხორციელდება:

- **ობიექტთა კლასიფიკაცია მოცემული ნიშან-თვისებების მიხედვით.**

  მაგალითად, შესაძლებელია, გამოვყოთ ამომრჩეველთა სხვადასხვა კლასტერი მათი ასაკის, საცხოვრებელი ადგილის, საქმიანობის, ეკონომიკური მდგომარეობის და ა. შ. ნიშნ(ებ)ით.

- **სტატისტიკური ერთობლიობის სტრუქტურის შესახებ შემუშავებული ჰიპოთეზის შემოწმება.**

  მაგალითად, შესაძლებელია, შევამოწმოთ ჰიპოთეზა იმის შესახებ, რომ ამომრჩევლები **k** კლასტერად იყოფიან.

- **სტატისტიკური ერთობლიობის სტრუქტურის აღწერა/ანალიზი.**

შესაძლოა დავახასიათოთ, რამდენი კლასტერისაგან შედგება სტატისტიკური ერთობლიობა, რა მოცულობისაა (რამდენი ობიექტისაგან შედგება) თითოეული კლასტერი, განვსაზღვროთ ცალკეული ნიშან-თვისების დისკრიმინაციული (განმასხვავებელი) მნიშვნელობა კლასტერების მიმართ და ა. შ.

იმისათვის, რომ კლასტერული ანალიზის მეშვეობით მონაცემთა სტრუქტურის **ოპტიმალური აღწერა** და **ინტერპრეტაცია** შევძლოთ, კლასტერიზაციის პროცესი საკვლევ ობიექტებს შორის არსებულ **მსგავსება-განსხვავებათა ადეკვატურად (შესაბამისად) უნდა წარიმართოს**.

თუ ობიექტებს/მოცემული სტატისტიკური ერთობლიობის წევრებს **საკოორდინატო სისტემაში განლაგებული წერტილების** სახით წარმოვიდგენთ, მაშინ:

[1] ინგლ.: cluster – კონა, ჯგუფი, შენადედი.

- ობიექტების **მსგავსება-განსხვავების** შესახებ შესაძლებელია ვიმსჯელოთ **წერტილთა შორის მანძილების**, ანუ იმის მიხედვით, რამდენად ახლოსაა/შორსაა წერტილები ერთმანეთისაგან.
- ობიექტებსა და წერტილებს შორის იგულისხმება შემდეგი სახის შესაბამისობა: რაც უფრო **მსგავსია** ობიექტები, მით **მცირეა მანძილი** საკოორდინატო სისტემაში მათ შესატყვის წერტილებს შორის.
- საკოორდინატო სისტემაში **კლასტერები წერტილთა არაცარიელი სიმრავლეების სახით იქნება წარმოდგენილი**.

## კლასტერიზაციის მეთოდები

**კლასტერიზაცია სხვადასხვა მეთოდის** მეშვეობით შეიძლება განხორციელდეს. ეს მეთოდები ორ ძირითად ჯგუფად იყოფა: **იერარქიულად** და **არაიერარქიულად**.

იერარქიული კლასტერიზაციის მეთოდის გამოყენების შედეგად მიიღება მონაცემთა იერარქიული, "ხისებრი" სტრუქტურა.

იერარქიული მეთოდები, თავის მხრივ, იყოფა **გამაერთიანებელ/აგლომერაციულ** და **დამყოფ/დივიზიურ** მეთოდებად.

გამაერთიანებელი/აგლომერაციული მეთოდების გამოყენებისას, კლასტერიზაციის საწყის ეტაპზე ითვლება, რომ თითოეული ობიექტი ცალკე კლასტერია და ხდება კლასტერების გაერთიანება, ანუ ობიექტების უფროდაუფრო დიდი მოცულობის ჯგუფებში თავმოყრა მანამ, სანამ საბოლოოდ ყველა ობიექტი ერთი კლასტერში არ აღმოჩნდება.

დამყოფი/დივიზიური მეთოდების გამოყენებისას, კლასტერიზაციის საწყის ეტაპზე ითვლება, რომ ყველა ობიექტი ერთ კლასტერში შედის და შემდეგ ხდება მისგან მცირე კლასტერების გამოყოფა, ანუ ობიექტების განაწილება უფროდაუფრო პატარა მოცულობის ჯგუფებში მანამ, სანამ საბოლოოდ სათითაო ობიექტი ცალკე კლასტერად არ გამოიყოფა.

ემპირიული სოციოლოგიური კვლევებისას, იერარქიული კლასტერიზაციის გამოყენების დროს ***უფრო ხშირად აგლომერაციულ მეთოდებს მიმართავენ.***

არაიერარქიული კლასტერიზაცია/**k**-საშუალოთა მეთოდი გულისხმობს თავდაპირველად **კლასტერის ცენტრის**, ანუ საწყისი წერტილის განსაზღვრას/შერჩევას, ხოლო შემდეგ ობიექტთა თავმოყრას ამ ცენტრის ირგვლივ – მათ დაჯგუფებას ცენტრიდან ზღვრული დაშორების/მანძილის მიხედვით.

არაიერარქიული კლასტერიზაციის მეთოდებია:

**თანმიმდევრული ზღვრული/"ზღურბლის" მეთოდი**. ამ მეთოდის გამოყენებისას ხდება კლასტერის ცენტრის შერჩევა და ობიექტების გაერთიანება ცენტრის ირგვლივ ზღვრული დაშორების/მანძილის მიხედვით.

**პარალელური ზღვრული/"ზღურბლის" მეთოდი.** ამ მეთოდის გამოყენებისას ხდება ერთდროულად/პარალელურად რამდენიმე კლასტერის ცენტრის შერჩევა და ობიექტების გაერთიანება ცენტრების ირგვლივ ზღვრული დაშორების/მანძილის მიხედვით.

**ოპტიმალური განაწილების მეთოდი**. ამ მეთოდის გამოყენებისას ობიექტები გადანაწილდება ისეთი წესით, რომ მოხდეს კლასტერებისშიდა და კლასტერებშორისი დაშორებებების/მანძილების ჯამური მაჩვენებლების ოპტიმიზება.

## აგლომერაციული იერარქიული კლასტერიზაციის მექანიზმი

აგლომერაციული იერარქიული კლასტერიზაციის მექანიზმი შემდეგნაირია:

- კლასტერიზაციის დაწყებამდე, მოცემული სტატისტიკური ერთობლიობის ყოველი წევრი/ობიექტი ითვლება ცალკე კლასტერად. ე. ი. თუ სტატისტიკურ ერთობლიობაში შედის **n** ობიექტი, ითვლება, რომ გვაქვს **n** კლასტერი.
- თავდაპირველად ხდება **ორი უახლოესი კლასტერის ამორჩევა და მათი ერთ კლასტერად გაერთიანება**. შედეგად, კლასტერების რაოდენობა ერთით მოიკლებს, ე. ი. დაგვრჩება **n-1** კლასტერი.
- შემდეგ **n-1** კლასტერს შორის ხდება ორი უახლოესი კლასტერის ამორჩევა და მათი გაერთიანება, მივიღებთ **n-2** კლასტერს. და ა. შ.
- იერარქიული კლასტერიზაციის პროცესი მთავრდება **ერთი** კლასტერის მიღებით, მაგრამ იგი შეიძლება შეწყდეს **ნებისმიერ** ეტაპზე, როდესაც რაიმე **კრიტერიუმის** მიხედვით ჩაითვლება, რომ მნიშვნელოვანი კლასტერები უკვე გამოყოფილია.

მაგალითად, ასეთი კრიტერიუმი შეიძლება იყოს **კლასტერთა ურთიერთდაშორების/კლასტერებს შორის მანძილის ზღვრულად მიჩნეული სიდიდე.** ე. ი. კლასტერების გაერთიანების პროცესი შეიძლება შევწყვიტოთ მაშინ, როდესაც

ორი უახლოესი კლასტერის ურთიერთდაშორება გარკვეულ **ზღვარს** გადააჭარბებს.

კლასტერთა გაერთიანებას **აგრეგაცია** ეწოდება.

აგლომერაციული იერარქიული კლასტერიზაციის შედეგად მიღებული, ეგრეთწოდებული "**იერარქიული ხე"** მოცემული სტატისტიკური ერთობლიობის **სტრუქტურას** გამოხატავს.

## კლასტერთა ურთიერთდაშორების ზომები

როგორც უკვე აღვნიშნეთ, საკოორდინატო სისტემაში კლასტერები წერტილთა არაცარიელი სიმრავლეების სახით შეიძლება წარმოვადგინოთ.

საკოორდინატო სისტემაში წერტილთა ურთიერთდაშორების ზომის/წერტილთა შორის მანძილის განსაზღვრა სხვადასხვა წესით/მეთოდით შეიძლება განხორციელდეს.

ასეთი მეთოდების მაგალითებია:

- **ევკლიდეს მანძილი.** ევკლიდეს მანძილი ეწოდება წერტილთა **გეომეტრიულ** დაშორებას მრავალგანზომილებიან საკოორდინატო სისტემაში და გამოითვლება ფორმულით:

$$\text{მანძილი } (\mathbf{x},\mathbf{y}) = \sqrt{\Sigma_i(\mathbf{x_i}-\mathbf{y_i})^2}$$

- **ევკლიდეს მანძილის კვადრატი.** ევკლიდეს მანძილის კვადრატი მიიღება ევკლიდეს მანძილის კვადრატში აყვანით, რაც დაშორების/მანძილის დატვირთვას/მნიშვნელობას ზრდის:

$$\text{მანძილი } (\mathbf{x},\mathbf{y}) = \Sigma_i(\mathbf{x_i}-\mathbf{y_i})^2$$

- **ხარისხობრივი მანძილი.** ხარისხობრივი მანძილი გამოითვლება ფორმულით:

$$\text{მანძილი } (\mathbf{x},\mathbf{y}) = \Sigma_i \left(\left| \mathbf{x_i}-\mathbf{y_i} \right|^{\mathbf{p}}\right)^{1/\mathbf{r}}$$

**P** და **r** პარამეტრების/ხარისხების იდენტიფიცირება ხდება მკვლევრის მიერ.

როდესაც **P = r =**2, ხარისხობრივი მანძილი ევკლიდეს მანძილს ემთხვევა.

ხარისხის მაჩვენებელთა ცვლილება საშუალებას გვაძლევს მოვახდინოთ დაშორების დატვირთვის/მნიშვნელობის ვარირება.

- **მანჰეტენის მანძილი.** მანჰეტენის მანძილი გამოითვლება ფორმულით:

$$\text{მანძილი } (\mathbf{x,y}) = \Sigma_i \left| \mathbf{x_i - y_i} \right|^2$$

- **ჩებიშევის მანძილი.** ჩებიშევის მანძილი გამოიყენება მაშინ, როდესაც საჭიროა განისაზღვროს ორ ობიექტს შორის განსხვავება ერთი კოორდინატის (ერთი განზომილების/ნიშან-თვისების მიხედვით):

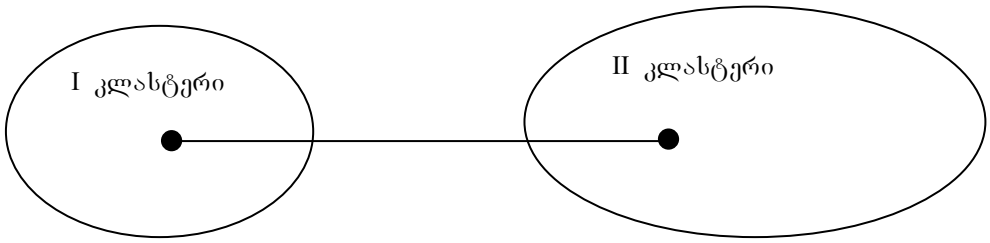
$$\text{მანძილი } (\mathbf{x,y}) = \mathbf{maximum} \left| \mathbf{x_i - y_i} \right|^2$$

- **უთავსობის პროცენტი.** გამოითვლება შემდეგი ფორმულით:

$$\text{მანძილი } (\mathbf{x,y}) = (\text{რაოდენობა } \mathbf{x_i \neq y_i)/i}$$

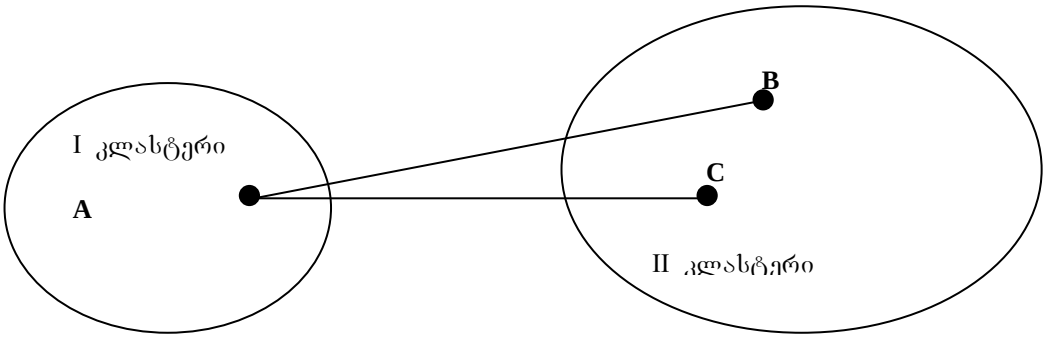
## კლასტერების აგრეგაცია

კლასტერიზაციის საწყის ეტაპზე, როდესაც ერთი კლასტერი ერთი ობიექტით, ე. ი. საკოორდინატო სისტემაში ერთი წერტილითაა წარმოდგენილი, კლასტერების აგრეგაცია რთული არაა, რადგან ორ კლასტერს შორის მანძილი ემთხვევა მანძილს/დაშორებას შესაბამის ორ წერტილს შორის:




ასეთ შემთხვევაში, როგორც ზემოთ აღვნიშნეთ, აგრეგაციის კრიტერიუმი წერტილთა წყვილებს შორის უმცირესი მანძილია.

მაგრამ სიტუაცია რთულდება, როდესაც ერთ-ერთ კლასტერში მაინც ერთზე მეტი წერტილი შედის:




ამ შემთხვევაში კლასტერებში შემავალ წერტილთა წყვილებს შორის სხვადასხვა მანძილია. რომელი მათგანი შეიძლება ჩაითვალოს მოცემულ კლასტერებს შორის მანძილად? **რა კრიტერიუმით გავაერთიანოთ კლასტერები?**

ეს პრობლემა რამდენიმე მეთოდით შეიძლება გადაიჭრას.

ასეთი მეთოდების მაგალითებია:

- **უახლოესი მეზობლების მეთოდი.** ორ კლასტერს შორის მანძილად მიჩნეულია მანძილი/დაშორება მათ უახლოეს წერტილებს შორის (ჩვენი მაგალითის შემთხვევაში, უახლოესი წერტილებია **A** და **C**). შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი უახლოეს წერტილებს/"მეზობლებს" შორის.
- **უშორესი მეზობლების მეთოდი.** ორ კლასტერს შორის მანძილად მიჩნეულია მანძილი/დაშორება მათ უშორეს წერტილებს შორის (ჩვენი მაგალითის შემთხვევაში, უშორესი წერტილებია **A** და **B**). შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი/დაშორება უშორეს წერტილებს/"მეზობლებს" შორის.
- **აუწონავი წყვილური საშუალოს მეთოდი.** ორ კლასტერს შორის მანძილი გამოითვლება როგორც წერტილების ყველა წყვილს შორის არსებულ მანძილთა/დაშორებათა საშუალო. შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია **უმცირესი საშუალო მანძილი/დაშორება**.
- **აწონილი წყვილური საშუალო.** კლასტერებს შორის მანძილი გამოითვლება აუწონავი წყვილური საშუალოს ანალოგიურად, მაგრამ გამოთვლებისას *თითოეული კლასტერის მოცულობა (ე. ი.*

*<u>კლასტერში შემავალი ობიექტების/წერტილების რაოდენობა) განიხილება როგორც წონითი კოეფიციენტი.</u>*

- **აუწონავი ცენტროიდული მეთოდი.** კლასტერებს შორის მანძილი განისაზღვრება როგორც მანძილი/დაშორება მათ ცენტრებს შორის. შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი/დაშორება კლასტერების ცენტრებს შორის.
- **აწონილი ცენტროიდული მეთოდი (medიანა).** აუწონავი ცენტროიდული მეთოდის ანალოგიურია, მაგრამ გაერთიანებული კლასტერის ცენტრის გამოთვლისას *თითოეული კლასტერის მოცულობა განიხილება როგორც წონითი კოეფიციენტი.*
- დისპერსიული ანალიზის მეთოდების გამოყენებით და ა. შ.

კლასტერთა ურთიერთდაშორების/კლასტერებს შორის მანძილის ზომები და აგრეგაციის მეთოდები ერთმანეთს უნდა შეესაბამებოდეს.

მაგალითად, თუ აგრეგაციას ცენტროიდული მეთოდის მეშვეობით ვაწარმოებთ, კლასტერთა ურთიერთდაშორება ევკლიდეს მანძილის კვადრატით უნდა გავზომოთ.

## ცვლადების სტანდარტიზება

კლასტერიზაცია რამდენიმე ნიშან-თვისების/ცვლადის მიხედვით იმ სირთულესთანაც არის დაკავშირებული, რომ საწყისი ცვლადების მნიშვნელობები სხვადასხვა ერთეულით შეიძლება იყოს წარმოდგენილი და ნებისმიერად შეიძლება იყოს "გაფანტული", რაც შედეგების ანალიზისა და ინტერპრეტაციის სირთულეს იწვევს. ეს პრობლემა ცვლადების სტანდარტიზებით წყდება.

კლასტერული ანალიზის დროს ცვლადების სტანდარტიზება სხვადასხვა მეთოდითაა შესაძლებელი:

- **Z-მნიშვნელობები**: Z-მნიშვნელობების მეთოდის გამოყენება გულისხმობს თითოეული ცვლადის საშუალო მნიშვნელობის გამოთვლასა და ამ მნიშვნელობის გაყოფას სტანდარტულ გადახრაზე.
- **ნორმირება (-1; 1)**: ნორმირება –1-დან 1-ის ფარგლებში გულისხმობს წრფივი გარდაქმნების მეშვეობით ცვლადების მნიშვნელობების მოქცევას **–1-სა და 1-ს შორის**.

- **ნორმირება (0; 1)**: ნორმირება 0-დან 1-ის ფარგლებში გულისხმობს წრფივი გარდაქმნების მეშვეობით ცვლადების მნიშვნელობების მოქცევას **0-სა და 1-ს შორის**.
- **მაქსიმუმზე გაყოფა**: მაქსიმუმზე გაყოფის მეთოდი გულისხმობს ცვლადების მნიშვნელობების გაყოფას მათ მაქსიმალურ მნიშვნელობაზე.
- **საშუალოზე გაყოფა**: ცვლადების მნიშვნელობები იყოფა მათ საშუალო მნიშვნელობაზე.
- **სტანდარტული გადახრაზე გაყოფა** და ა.შ.

## არაიერარქიული კლასტერიზაციის მექანიზმი

## ოპტიმალური განაწილების მეთოდი

**K საშუალოთა მეთოდი** ძირითადად მაშინ გამოიყენება, როდესაც ხდება **ჰიპოთეზის შემოწმება სტატისტიკური ერთობლიობის სტრუქტურის** შესახებ.

დავუშვათ, შემუშავებულია ჰიპოთეზა, რომლის მიხედვითაც, სტატისტიკური ერთობლიობის სტრუქტურა **k** კლასტერისაგან შედგება.

**K საშუალოთა მეთოდის** მეშვეობით შესაძლებელია სტატისტიკური ერთობლიობა დავყოთ ზუსტად **k** კლასტერად, რომლებიც **ერთმანეთისაგან რაც შეიძლება მეტადაა დაშორებული/რომელთა შორისაც რაც შეიძლება მეტი მანძილია.**

K საშუალოთა მეთოდით კლასტერიზაციის მექანიზმი შემდეგნაირია:

- თავდაპირველად ხდება **k** კლასტერის **შემთხვევითი** შერჩევა.
- შემდეგ კლასტერებს შორის ობიექტები **გადანაცვლდება/გადანაწილდება** იმ წესით, რომ მოხდეს:

ა) **კლასტერებისშიდა მანძილების ვარიაციათა მინიმიზება.**

ბ) **კლასტერებსშორისი მანძილების ვარიაციათა მაქსიმიზება.**

**K საშუალოთა მეთოდით** კლასტერების გამოყოფის შემდეგ უნდა შეფასდეს **რამდენად მნიშვნელოვანია განსხვავება/დისკრიმინაცია მიღებულ კლასტერებს** შორის.

კლასტერებს შორის განსხვავების **სტატისტიკური მნიშვნელოვნება** შეიძლება შეფასდეს **ფიშერის კრიტერიუმის** გამოყენებით კლასტერებისშიდა და კლასტერებშორისი მანძილების ვარიაციების შედარებითი ანალიზის საფუძველზე, ან **T**-კრიტერიუმის (სტიუდენტის განაწილების) მეშვეობით: **კლასტერული ცენტროიდების[2] შედარებითი ანალიზის** საფუძველზე.

**თუ კლასტერებს შორის განსხვავება შეფასდება როგორც მნიშვნელოვანი, ჰიპოთეზა სტატისტიკური ერთობლიობის k კლასტერად დაყოფის შესახებ ვერიფიცირებულად ჩაითვლება, ხოლო წინააღმდეგ შემთხვევაში – ფალსიფიცირებულად.**

## კლასტერიზაციის მეთოდის შერჩევის პრობლემა

კლასტერიზაციის სხვადასხვა მეთოდის არსებობა, ბუნებრივია, ბადებს კითხვას: რის მიხედვით შევარჩიოთ კლასტერიზაციის კონკრეტული მეთოდი მონაცემთა ანალიზისას?

ამ კითხვაზე ერთმნიშვნელოვანი პასუხის გაცემა შეუძლებელია. გადაწყვეტილება მკვლევარმა უნდა მიიღოს კვლევის მიზნის/ამოცანებისა და კლასტერიზაციის სხვადასხვა მეთოდის სპეციფიკის მიხედვით, მაგრამ მიზანშეწონილია შემდეგი, ზოგადი სახის რეკომენდაციების გათვალისწინება:

- იერარქიული კლასტერიზაციის გამოყენება უფრო პროდუქტიულია მაშინ, **როდესაც სტატისტიკური ერთობლიობის სტრუქტურა უცნობია.** ამ შემთხვევაში, კლასტერიზაცია ერთგვარი "დაზვერვითი" საშუალებაა იმ თვალსაზრისით, რომ ხდება მონაცემთა სტრუქტურის გამოვლენა, რის საფუძველზეც შესაძლებელია შემდგომი ანალიზის სწორი მიმართულებით წარმართვა/კორექტული დაგეგმვა.
- მონაცემთა სტრუქტურის შესახებ არსებული **ჰიპოთეზის შემოწმების ოპტიმალური საშუალება არაიერაქიული კლასტერიზაციაა *(ჰიპოთეზის ფორმულირების შესაძლებლობა უკვე ინფორმირებულობის გარკვეულ ხარისხზე მეტყველებს).***
- არაიერარქიული კლასტერიზაციის მეთოდები იერარქიულთან შედარებით უფრო სწრაფია და მისი გამოყენება მოხერხებულია ობიექტთა დიდი რაოდენობის მიმართ.

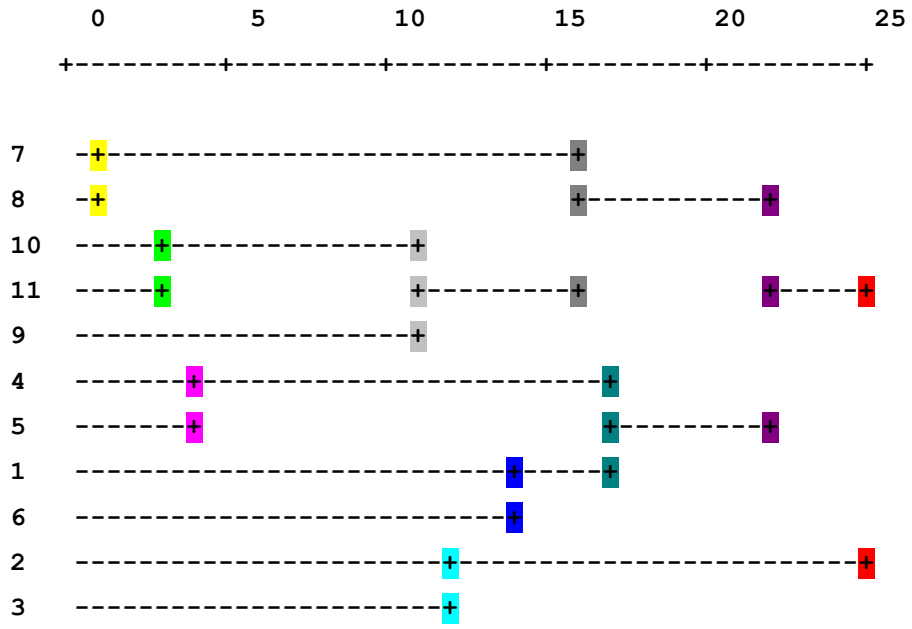
[2] **კლასტერული ცენტროიდი ეწოდება განზომილების/ნიშან-თვისების საშუალო მნიშვნელობას მოცემული კლასტერის ფარგლებში.**

- რიგ შემთხვევაში, მიზანშეწონილია იერარქიული და არაიერაქიული მეთოდების ერთობლივი გამოყენება: თავდაპირველად, იერარქიული მეთოდის მეშვეობით შეიძლება განისაზღვროს კლასტერების რაოდენობა, შემდეგ კი, არაიერაქიული მეთოდით მოხდეს დადგენილი რაოდენობის კლასტერებად ობიექტთა ოპტიმალური განაწილება.

## დენდროგრამა

დენდროგრამა, ანუ ეგრეთწოდებული იერარქიული ხე კლასტერიზაციის შედეგების გრაფიკული გამოსახვის/ვიზუალიზაციის ერთ-ერთი გავრცელებული მეთოდია.

დენფროგრამის ზოგადი სახე შემდეგნაირია:




- დენდროგრამის გარჩევა (გაშიფვრა, ”წაკითხვა”) ხდება მარცხნიდან მარჯვნივ. + -ებს შორის არსებული ვერტიკალური დაშორებების მიხედვით შესაძლებელია ვიმსჯელოთ იმის შესახებ, თუ როგორ მიმდინარეობს აგრეგაციის პროცესი (როგორ ერთიანდება კლასტერები): დავადგინოთ, რომელი კლასტერები ერთიანდება თითოეულ მიღებულ კლასტერში, განვსაზღვროთ კლასტერების რაოდენობა და მათი აგრეგაციის მიმდევრობა.
- კლასტერების ზემოთ მოთავსებული ჰორიზონტალური მონაკვეთი დაშორებების/მანძილების ზომებს გამოსახავს. შესაბამისად, მასზე

მითითებული ინტერვალების მეშვეობით შესაძლებელია განვსაზღვროთ რა დაშორების მიხედვით/რა მანძილზე მოხდა კლასტერთა აგრეგაცია.

- ეს ინფორმაცია შესაძლებელია გამოვიყენოთ როგორც წანამძღვარი გადაწყვეტილების მისაღებად იმის შესახებ, თუ როდის დავასრულოთ კლასტერიზაციის პროცესი, ანუ **რამდენი კლასტერი გამოვყოთ** იმისათვის, რომ მოცემული სტატისტიკური ერთობლიობის სტრუქტურა რაც შეიძლება ოპტიმალურად აღიწეროს: კლასტერიზაცია შეიძლება დასრულებულად ჩავთვალოთ იმ ეტაპზე, რომლის შემდეგაც **მკვეთრად იზრდება მანძილი/დაშორება კლასტერებს შორის**.
- დენდროგრამის მეშვეობით აგრეთვე ვიღებთ თვალსაჩინო ინფორმაციას იმის შესახებ, ეკუთვნის თუ არა მოცემულ კლასტერს სტატისტიკური ერთობლიობის ესა თუ ის წევრი/ობიექტი.