

Informatikos fakultetas

# **P176B101 Intelektikos pagrindai**

**komandinio darbo ataskaita**

|  |  |  |
| --- | --- | --- |
|  | Studentas: | Ignas Šakys, IFF-7/5 |
|  |  |
| Dėstytoja: | Doc. Agnė Paulauskaitė - Tarasevičienė |

Kaunas 2021

Turinys

[**P176B101 Intelektikos pagrindai** 1](#_Toc72250911)

[1. Paveiksliukų sąrašas 3](#_Toc72250912)

[2. Naudotas duomenų rinkinys 4](#_Toc72250913)

[2.1. Tolydinio tipo kintamieji 4](#_Toc72250914)

[2.2. Kategorinio tipo kintamieji 4](#_Toc72250915)

[3. SOM realizuojantys metodai 5](#_Toc72250916)

[4. Atstumo metrikos 5](#_Toc72250917)

[5. Atlikti eksperimentai 6](#_Toc72250918)

[6. Išvados 10](#_Toc72250919)

# 

# Paveiksliukų sąrašas

[pav. 1 Tolydinio tipo kintamųjų analizės lentelė prieš duomenų koregavimą 4](#_Toc72250921)

[pav. 2 Tolydinio tipo kintamųjų analizės lentelė po duomenų koregavimo 4](#_Toc72250922)

[pav. 3 Kategorinio tipo kintamųjų analizės lentelė prieš duomenų koregavimą 4](#_Toc72250923)

[pav. 4 Kategorinio tipo kintamųjų analizės lentelė po duomenų koregavimo 4](#_Toc72250924)

[pav. 5 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 3 6](#_Toc72250925)

[pav. 6 Antrojo bandymo SOM žemėlapio grafikas, kai k = 3 6](#_Toc72250926)

[pav. 7 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 4 7](#_Toc72250927)

[pav. 8 Antrojo bandymo SOM žemėlapio grafikas, kai k = 4 7](#_Toc72250928)

[pav. 9 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 5 8](#_Toc72250929)

[pav. 10 Antrojo bandymo SOM žemėlapio grafikas, kai k = 5 8](#_Toc72250930)

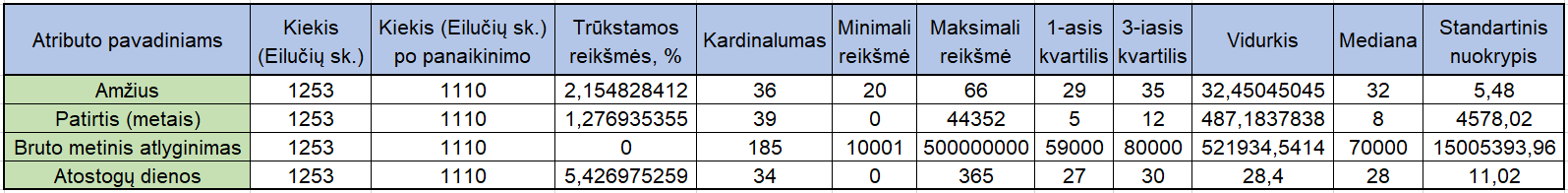
[pav. 11 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 6 9](#_Toc72250931)

[pav. 12 Antrojo bandymo SOM žemėlapio grafikas, kai k = 6 9](#_Toc72250932)

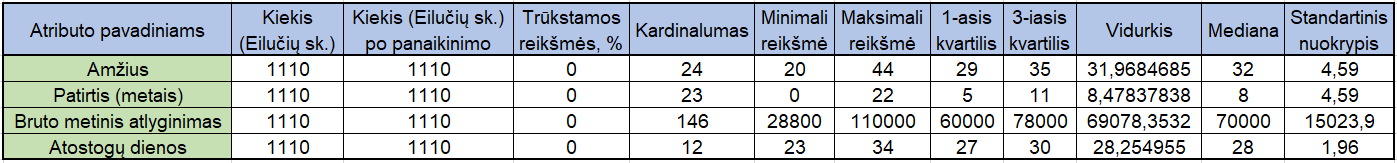
# Naudotas duomenų rinkinys

Darbui atlikti buvo pasirinktas pirmojo laboratorinio darbo metu naudotas duomenų rinkinys. Šiam rinkiniui atlikus duomenų analizę matome, jog dėl tuščių reikšmių laukuose, buvo pašalinta 143 eilutės ir duomenų rinkinyje liko 1110 eilučių su kuriomis dirbsime toliau. Žvelgiant į kitus rodiklius galime pastebėti, kad duomenys yra iškraipyti dėl duomenyse esančių neadekvačių reikšmių, jas sutvarkysime šalindami duomenų kokybės problemas.

# Tolydinio tipo kintamieji

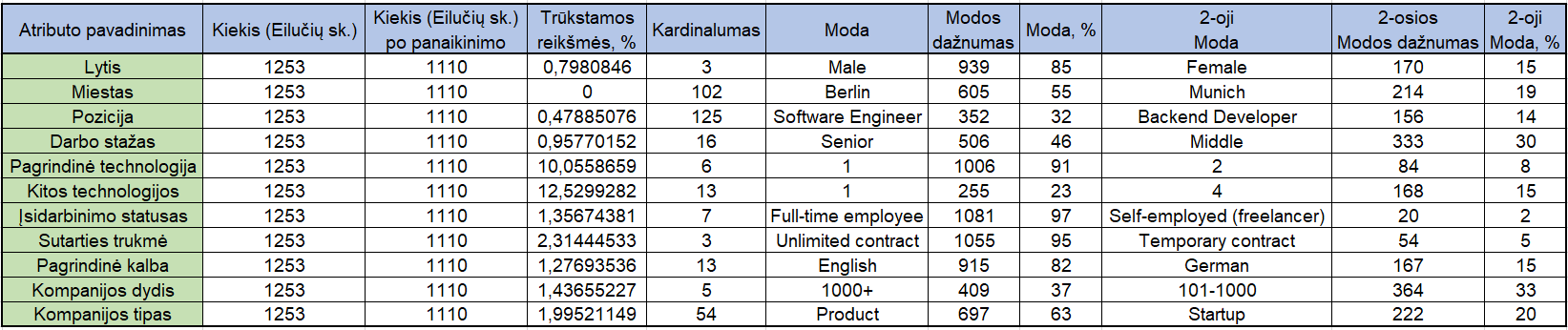


pav. 1 Tolydinio tipo kintamųjų analizės lentelė prieš duomenų koregavimą

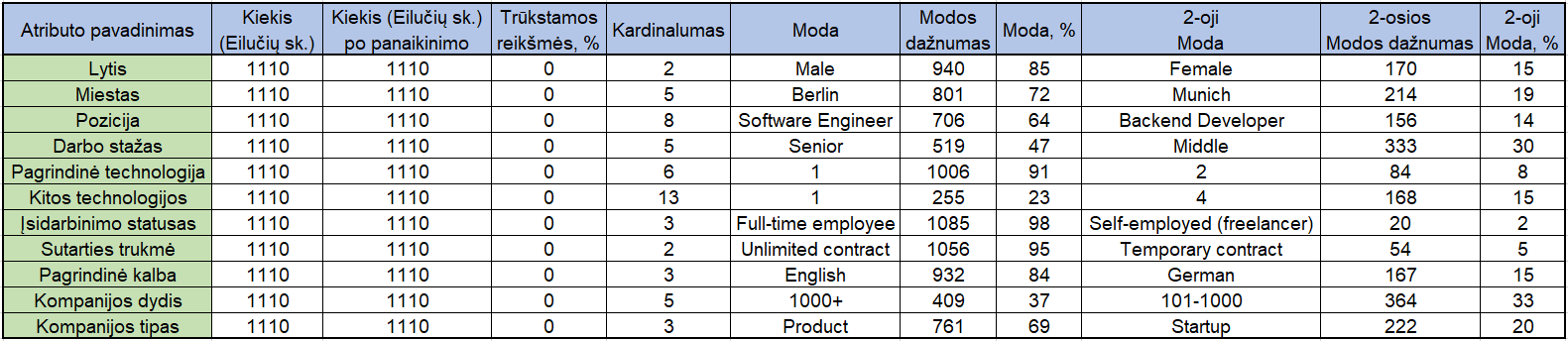


pav. 2 Tolydinio tipo kintamųjų analizės lentelė po duomenų koregavimo

# Kategorinio tipo kintamieji



pav. 3 Kategorinio tipo kintamųjų analizės lentelė prieš duomenų koregavimą



pav. 4 Kategorinio tipo kintamųjų analizės lentelė po duomenų koregavimo

# SOM realizuojantys metodai

Pirmiausia nuskaitomi ir sutvarkomi duomenys iš pasirinkto duomenų failo. Toliau inicijuojama SOM klasė ir kviečiamos šios klasės metodas “train()“. Galiausiai gautas SOM žemėlapis atvaizduojamas grafiškai.

Norint realizuoti SOM bus vykdomi sekantys žingsniai:

1. SOM modelio iniciavimas, kai modeliui paduodamos žemėlapio dimensijų dydžiai, mokymosi greitis, bei nuskaityti ir į duomenų rinkinį suformuoti duomenys. Programa pasinaudodama jai suteiktais kintamaisiais, sugeneruos žemėlapio duomenų struktūrą, kurioje bus inicijuojamos atsitiktinės vektorių reikšmės intervale nuo 0 iki 1.
2. Išrenkamas atsitiktinis vektorius iš įvesties duomenų rinkinio.
3. Skaičiuojamas BMU (*angl. Best Matching Unit*) arba kitaip dar vadinamas neuronas – nugalėtojas. Šio neurono vektorius yra labiausiai panašus į atsitiktinai parinkto vektoriaus iš įvesties duomenų. Neuronas – nugalėtojas išrenkamas naudojant atstumo tarp dviejų vektorių apskaičiavimo metodus, šiuo atveju metodas remiasi Euklido atstumo formule.
4. Žinant BMU yra apskaičiuojamas vadinamosios neurono kaimynystės dydis. Kitaip tariant surandami šalia šio neurono esantys neuronai.
5. BMU ir kaimynystei priklausančių neuronų svorių perskaičiavimas.
6. Mokymosi greičio mažinimas.
7. Žingsniai nuo 2 iki 6 vykdomi tol, kol suformuojamas galutinis SOM žemėlapis.

# Atstumo metrikos

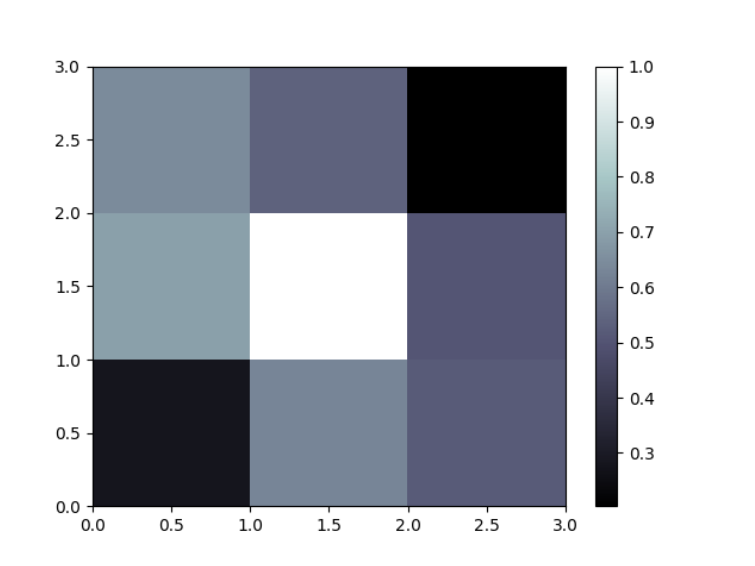
Kritinė SOM realizacijos dalis, kuria remiantis veikia save organizuojantys žemėlapiai – atstumo tarp atitinkamų taškų skaičiavimo metodas. Šio darbo realizacijoje naudotas Euklido atstumas, kitaip tariant atstumo tarp parinkto neurono ir įvesties vektoriaus kvadratas. Nors šiuo konkrečiu atveju Euklido atstumo skaičiavimo metodas ir atlieka skaičiavimus korektiškai, tačiau didėjant duomenų kiekiui, kai kalbama apie milžiniškus duomenų rinkinius, specialistai renkasi kitokius alternatyvius metodus.

# Atlikti eksperimentai

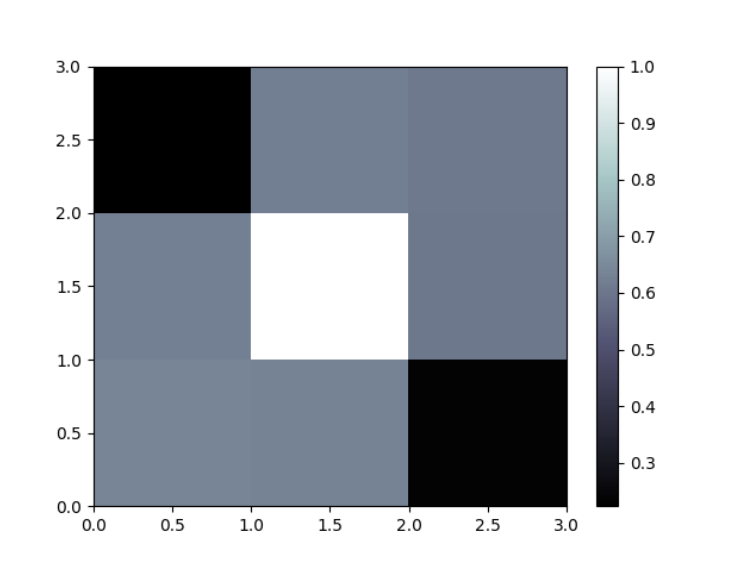
Prieš atliekant eksperimentus, buvo priimtos sekančių konstantų reikšmės, bei paaiškinimai:

* Mokymosi greičio pradinė reikšmė – 0.5;
* Epochų skaičius – 250;
* Kuo grafiko langelio spalva tamsesnė, tuo atstumas tarp neuronų yra mažesnis;
* Kuo grafiko langelio spalva šviesesnė, tuo atstumas tarp neuronų yra didesnis;
* Tamsesnių langelių susitelkimas atvaizduoja susiformavusius klasterius;
* Šviesesnių langelių susitelkimas atvaizduoja atskirtį tarp susiformavusių klasterių;

Pirmojo bandymo metu klasterių skaičius buvo lygus 3. Programa kompiliuota du kartus, gauti du skirtingi SOM žemėlapių grafikai pavaizduoti 5 ir 6 paveikslėliuose.

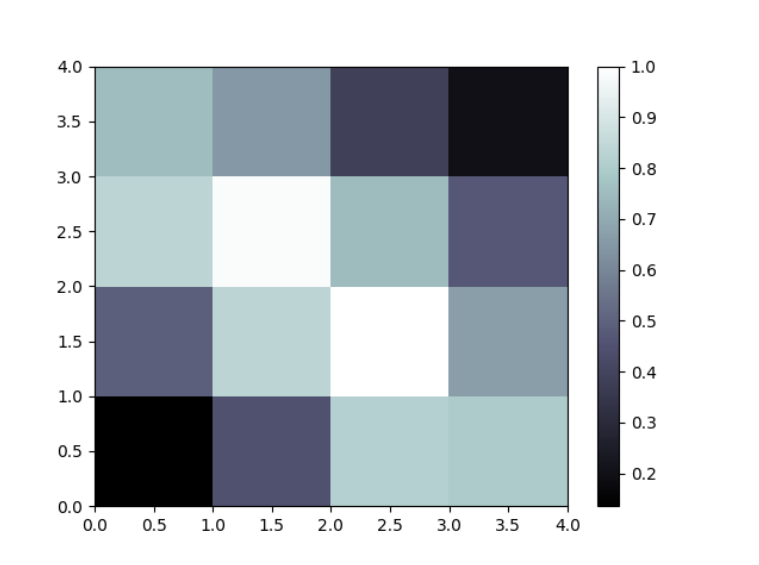


pav. 5 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 3

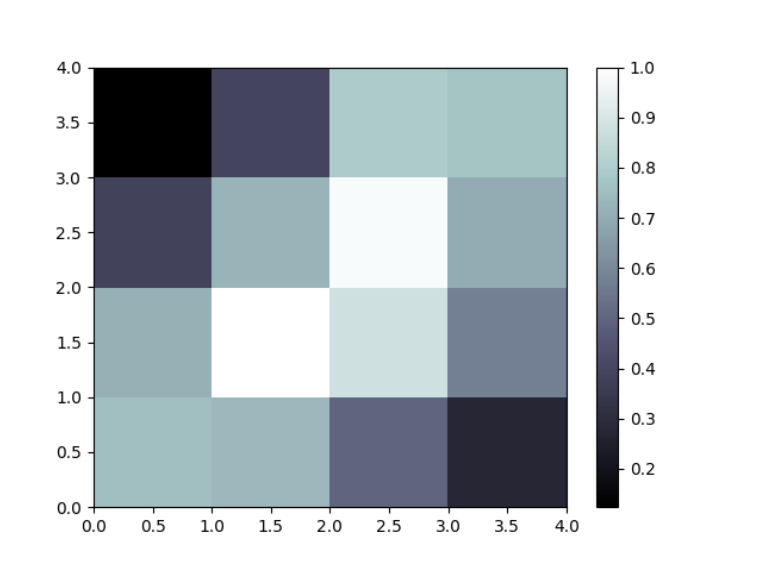


pav. 6 Antrojo bandymo SOM žemėlapio grafikas, kai k = 3

Antrojo bandymo metu klasterių skaičius buvo lygus 4. Programa kompiliuota du kartus, gauti du skirtingi SOM žemėlapių grafikai pavaizduoti 7 ir 8 paveikslėliuose.

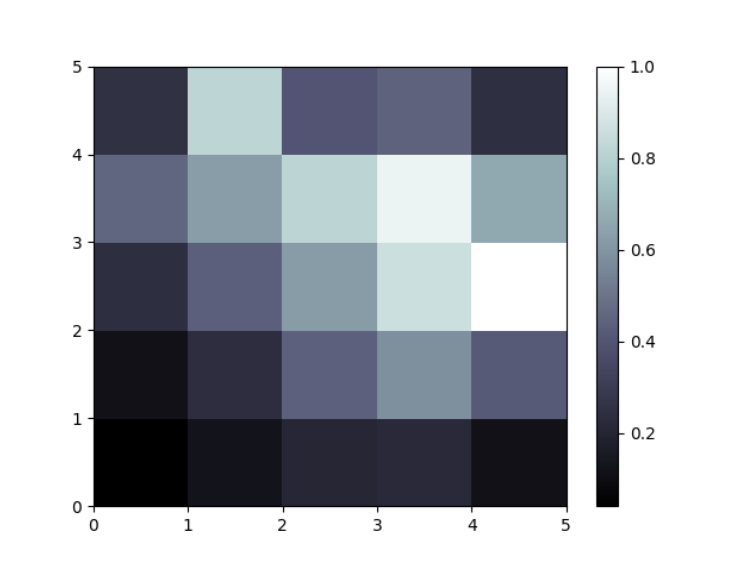


pav. 7 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 4

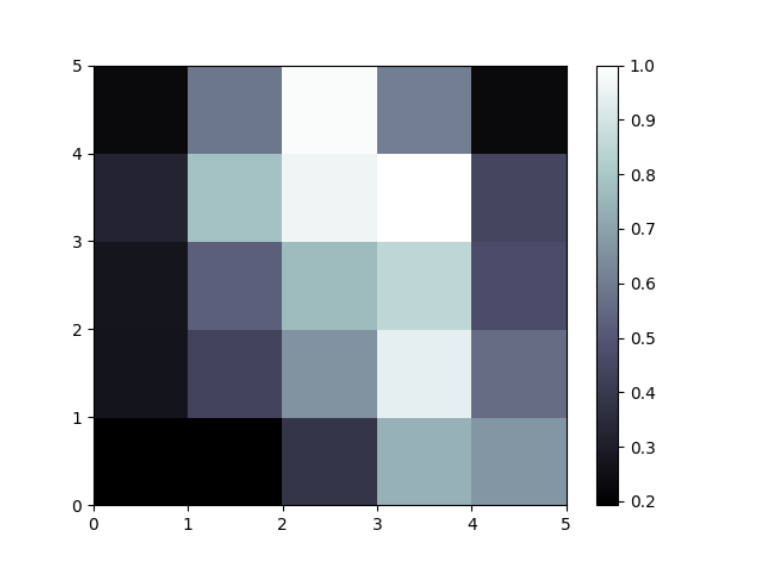


pav. 8 Antrojo bandymo SOM žemėlapio grafikas, kai k = 4

Trečiojo bandymo metu klasterių skaičius buvo lygus 5. Programa kompiliuota du kartus, gauti du skirtingi SOM žemėlapių grafikai pavaizduoti 9 ir 10 paveikslėliuose.

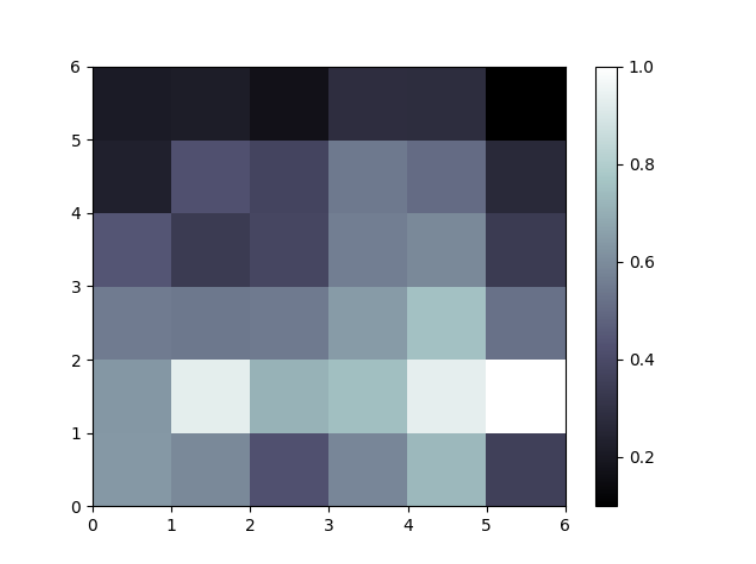


pav. 9 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 5

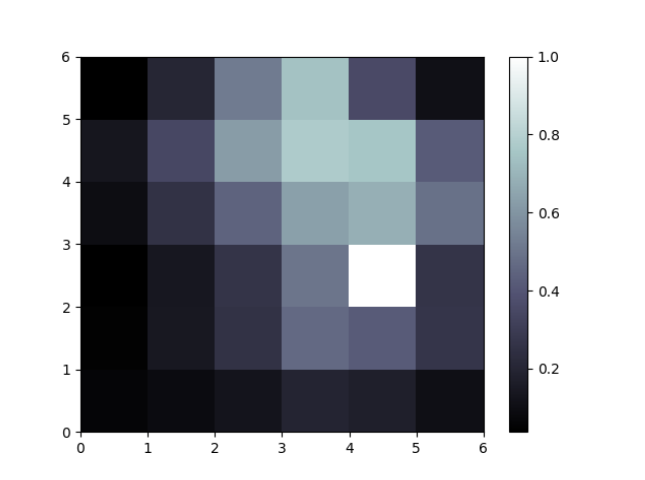


pav. 10 Antrojo bandymo SOM žemėlapio grafikas, kai k = 5

Ketvirtojo bandymo metu klasterių skaičius buvo lygus 5. Programa kompiliuota du kartus, gauti du skirtingi SOM žemėlapių grafikai pavaizduoti 11 ir 12 paveikslėliuose.



pav. 11 Pirmojo bandymo SOM žemėlapio grafikas, kai k = 6



pav. 12 Antrojo bandymo SOM žemėlapio grafikas, kai k = 6

# Išvados

Atlikus SOM realizaciją ir išnagrinėjus grafiškai pateiktus rezultatus, kuriuose atitinkama spalva atvaizduojami svorių pasiskirstymai suformuotame neuronų žemėlapyje, padarytos sekančios išvados:

1. Gautuose SOM žemėlapių grafikuose, matomi sudaryti klasteriai, tačiau dėl pasirinkto duomenų rinkinio, ne visuose grafikuose matoma aiški klasterių atskirtis.
2. Euklido atstumo metodas tinkamas mažesniems duomenų rinkiniams, didesniuose duomenų rinkiniuose yra naudojami kitokie atstumo apskaičiavimo metodai.
3. Eksperimentų atlikimo eigoje, pastebėta, jog geresni klasterių suformavimo rezultatai gaunami naudojant 250 epochų. Panaudojus 100 epochų aptikta neapsimokymo situacija, t. y. per 100 epochų nebūdavo suformuojami tokie aiškūs klasteriai.