

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS
PROGRAMŲ SISTEMŲ STUDIJŲ PROGRAMA

Saviorganizuojantys neuroniniai tinklai

5 užduotis

Atliko: 4 kurso 1 grupės studentė
Rosita Raišutytė

TURINYS

1. ĮVADAS	2
1.1. Tikslas	2
1.2. Uždaviniai	2
2. TYRIMAS	3
2.1. Duomenys.....	3
2.2. Programos kodas	3
2.3. Hyperparametrai	3
3. REZULTATAI	4
3.1. SOM žemėlapis 5x5	4
3.2. SOM žemėlapis 10x10.....	5
3.3. Apibendrinimas	5
4. IŠVADOS	6

1. Įvadas

1.1. Tikslas

Užduoties tikslas – suprogramuoti saviorganizuojančio neuroninio tinklo (žemėlapių, SOM) mokymo algoritmą, apmokyti jį naudojant irisų duomenis ir gauti rezultatų žemėlapi.

1.2. Uždaviniai

- Parašyti kodą, kuris sukurtų, apmokytų ir ištestuotų SOM tinklą.
- Apmokyti SOM tinklą su irisų duomenų rinkiniu su skirtingo dydžio žemėlapiais: 5x5, 10x10.
- Ištestuoti apmokytą SOM tinklą, gauti kvantavimo paklaidą ir klasterizuotų duomenų žemėlapi.
- Pavaizduoti gautą rezultatų žemėlapi.

2. Tyrimas

2.1. Duomenys

Tyrimui naudojami duomenys buvo paimti iš "C Irvine Machine Learning Repository" puslapio. Tyrimui naudotas irisų (angl. iris) duomenų rinkinys.

Irisų duomenų rinkinys sudaro:

- 150 duomenų įrašų;
- 50 duomenų įrašų yra Setosa klasės;
- 50 duomenų įrašų yra Versicolor klasės;
- 50 duomenų įrašų yra Virginica klasės;
- Setosa klasė buvo pervadinta į 0;
- Versicolor klasė buvo pervadinta į 1;
- Virginica klasė buvo pervadinta į 2;
- duomenų įrašą sudaro 4 požymiai ir klasė.

Prieš pradėdant dirbti su duomenimis klasės buvo pervadintos į 0, 1 ar 2. Duomenų požymiai buvo normalizuoti (standartizuoti), kad būtų nuo 0 iki 1. SOM mokymui buvo naudojami tik duomenų įrašų požymiai, modeliui klasės buvo nežinomos.

2.2. Programos kodas

Programos kodą galite rasti paspaudę nuorodą [čia](#).

2.3. Hyperparametrai

Tyrimo metu naudoti hyperparametrai:

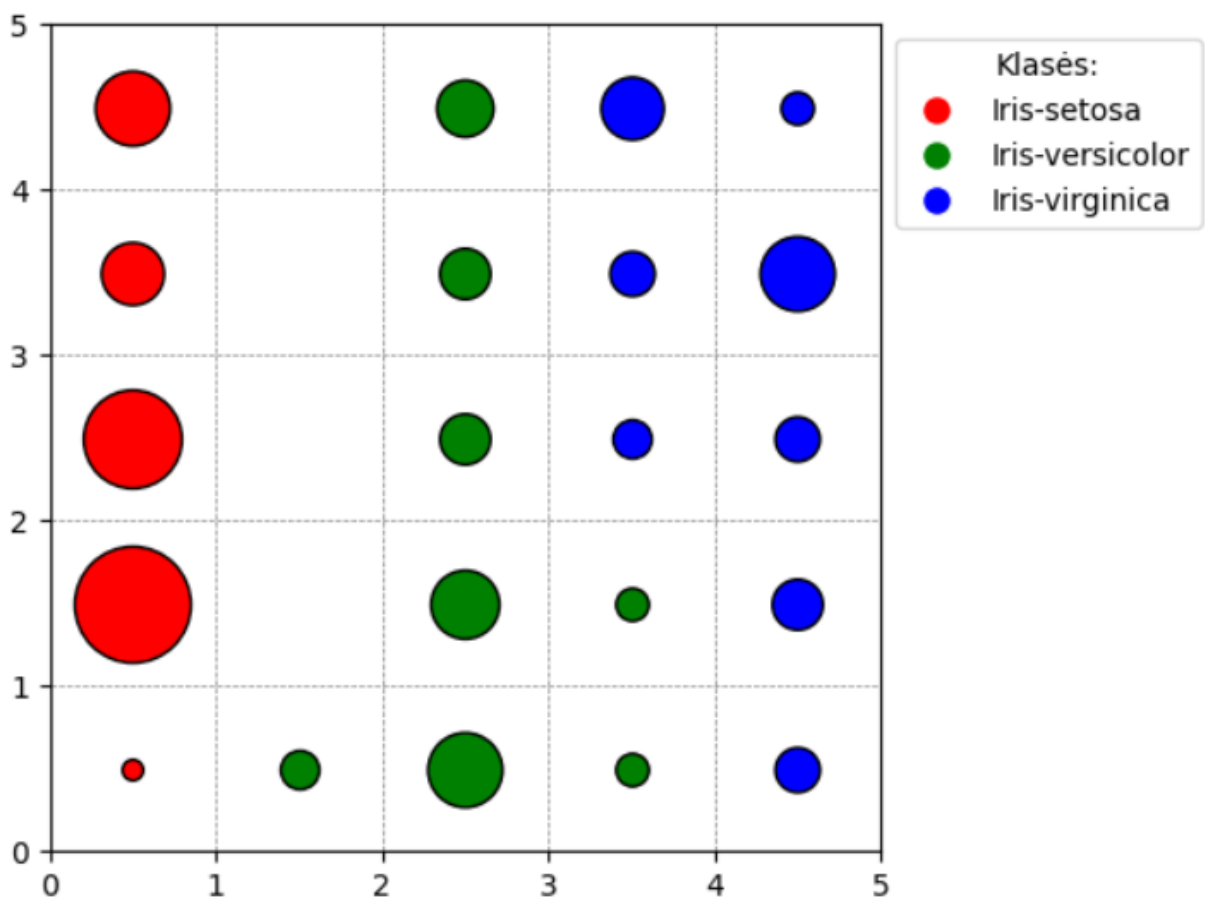
- SOM tinklo žemėlapių dydis: 5x5, 10x10;
- mokymosi epochų skaičius – 500;
- pradinis mokymosi parametras – 0,5;
- pradinis manhatteno atstumas – 4.

3. Rezultatai

Klasterizuotų duomenų žemėlapiai vaizduojami xy plokštumoje, kurioje uždėtas tinklelis skiria įsivaizduojamus SOM neuronus. Neuronai, kuriems buvo priskirti duomenų įrašai, vaizduojami su skrituliais. Skirtingos skritulių spalvos žymi skirtingas klases: raudona – Iris-Setosa, žalia – Iris-Versicolor, mėlyna – Iris-Virginica. Skirtingi dydžiai žymi tos klasės įrašų kiekį, kuo daugiau įrašų priskirta neuronui, tuo didesnis skritulys. Skritulio spalva ir dydis atrenkamas pagal laiminčiąją klasę tame neurone.

3.1. SOM žemėlapis 5x5

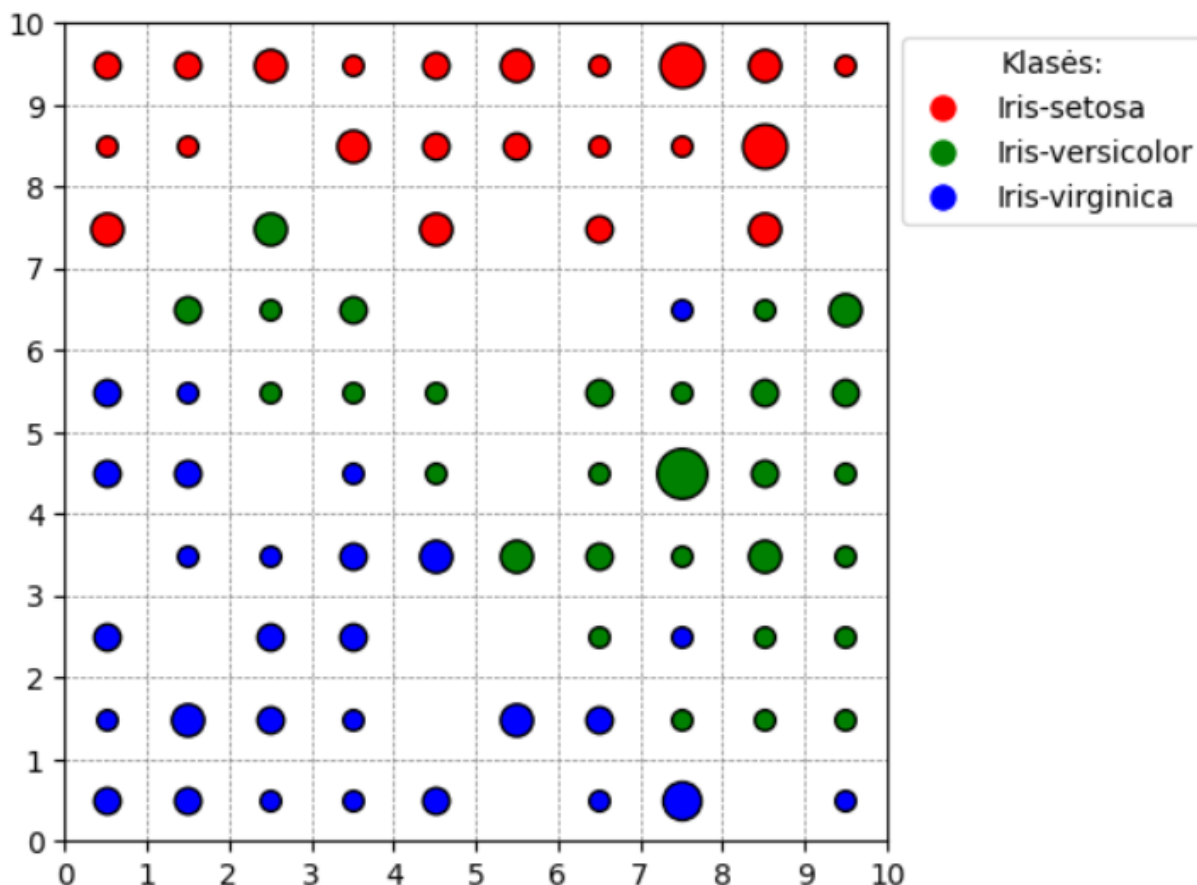
Sukūrus ir apmokius 5x5 dydžio SOM žemėlapi testavimo metu gauta 0,088 kvantavimo paklaida. 1 pav. matomi suklasterizuoti irisų duomenys. Matyti, kad Iris-Versicolor ir Iris-Virginica yra labai arti viena kitos ir be skirtingų spalvų žemėlapyje būtų galima manyti, kad tai viena klasė. Taip pat kadangi vaizduojama tik pirmaujanti klasė neurone, nėra aišku ar tame neurone nėra kitos klasės duomenų įrašų, kurių gali būti šiek tiek mažiau.



1 pav. 5x5 SOM žemėlapis klasterizavimo rezultatas

3.2. SOM žemėlapis 10x10

Sukūrus ir apmokius 10x10 dydžio SOM žemėlapi testavimo metu gauta 0,034 kvantavimo paklaida. 2 pav. matomi suklasterizuoti irisų duomenys. Matyti, kad Iris-Versicolor ir Iris-Virginica duomenys yra dar labiau persipynę ir susimaišę. Nors didesnis žemėlapis turėjo padėti atskirti klases, turimas didesnis neuronų kiekis leido duomenų įrašams labiau pasiskleisti ir prisitaikyti neuroną pagal savo klasę.



2 pav. 10x10 SOM žemėlapijo klasterizavimo rezultatas

3.3. Apibendrinimas

Iš abiejų rezultatų žemėlapių matyti, kad Iris-Versicolor ir Iris-Virginica klasės yra labai panašios savo požymiais, todėl jų klasteriai išsidėsto vieni šalia kitų ir perpina. Labiausiai išsiskiria Iris-Setosa klasė, ji šiek tiek artimesnė Iris-Versicolor, o labiausiai nutolusi nuo Iris-Virginica klasės.

Nors 2 pav. esančiame žemėlapyje klasių susimaišymas didesnis, tačiau kaip matyti 1 lentelėje kvantavimo paklaida mažesnė žemėlapiui 10x10.

1 lentelė. SOM tinklo kvantavimo paklaida pagal žemėlapių dydį

SOM žemėlapių dydis	Kvantavimo paklaida
5x5	0,088
10x10	0,034

4. Išvados

- SOM tinklas 10x10 geriau klasterizuoja irisų duomenis lyginant su 5x5 SOM žemėlapiu.
- Tyrimo metu pasirinktas toks SOM tinklo rezultatų žemėlapis nevisiškai atspindi klasių išsidėstymo, kadangi neurone atvaizduojama tik laiminčioji klasė.
- Irisų klasės Iris-Versicolor ir Iris-Virginica yra labai panašios savo požymiais, todėl SOM žemėlapyje išsidėsto šalia viena kitos ir perpina.