VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

PROGRAMŲ SISTEMOS

**Skaitmeninis intelektas ir sprendimų priėmimas**

**Tiesinio sklidimo DNT naudojant sistemą WEKA**

**Darbą atliko:**

Pijus Zlatkus

Vilnius

2024

Turinys

[1. Užduoties tikslas 3](#_Toc165146212)

[2. Duomenų paruošimas 3](#_Toc165146213)

[3. Pirma užduočių seka 4](#_Toc165146214)

[3.1. Neuroninio tinklo parametrų parinkimas 5](#_Toc165146215)

[4. Antra užduočių seka 10](#_Toc165146216)

[5. Trečia užduočių seka 13](#_Toc165146217)

[6. MS Excel dirbtinio neuroninio tinklo modelis 16](#_Toc165146218)

# Užduoties tikslas

Šios užduoties tikslas – išmokyti neuroninį tinklą teisingai klasifikuoti duomenis naudojant sistemą WEKA.

**Studento numeris:** 2110648

**Užduoties variantas:** sepallength, petallength, petalwidth

# Duomenų paruošimas

Šiai užduočiai atlikti buvo naudojamas Irisų duomenų rinkinys. Šis duomenys ne tik yra pasiekiami internetu, be yra *arff* failo formato WEKA programos failų aplinkoje kaip pavyzdinis duomenų rinkinys.

Šiai užduočiai atlikti, reikėjo iš vieno irisų duomenų rinkinio padaryti du. Kadangi šiame rinkinyje yra iš viso 150 įrašų, arba po 50 kiekvienai klasei), tai pirmame faile *iris\_train\_test.arff* reikėjo palikti 40 kiekvienos klasės įrašų, o *iris\_new.arff* tik po 10. *iris\_train\_test.arff* faile esantys duomenys bus naudojami neuroninio tinklo mokymui, o *iris\_new.arff* testavimui.

Darant užduočių sekas vienas iš reikalavimų buvo pašalinti vieną požymį ir palikti tik tuos kurie atitinka užduoties variantą - sepallength, petallength, petalwidth.

A screenshot of a computer program

Description automatically generated

**1 pav.** iris\_train\_test.arff failo pavyzdys

# Pirma užduočių seka

A diagram of a computer network

Description automatically generated

**2 pav.** Pirma užduočių seka

WEKA platformoje buvo sukurta užduoties seka (žr. **2 pav.** Pirma užduočių seka), kurioje yra apmokomas modelis. Įtraukti šie komponentai: ArffLoader, ClassAssigner, Remove, TextViewer, CrossValidationFoldMaker, MultilayerPerceptron, SerializedModelSaver ir ClassifierPerformanceEvaluator. Į ArffLoader komponentą įkeltas *iris\_train\_test.arff* duomenų rinkinys. Komponente Remove pašalintas antras požymis, kuris pagal užduotį nėra reikalingas neuroninio tinklo mokymui. Neuroninio tinklo modelis įrašomas į nurodytą kompiuterio aplanką per SerializedModelSaver. CrossValidationFoldMaker nustatytas penkių blokų kryžminės patikros skaičius, o MultilayerPerceptron komponente paketo dydis (BatchSize) pakeistas į 10.

## Neuroninio tinklo parametrų parinkimas

A screenshot of a computer program

Description automatically generated

**3 pav.** Pirmas parametrų parinkimo bandymas

A screenshot of a computer

Description automatically generated

**4 pav.** Antras parametrų parinkimo bandymas

A screenshot of a computer

Description automatically generated

**5 pav.** Trečias parametrų parinkimo bandymas

A screenshot of a computer program

Description automatically generated

**6 pav.** Ketvirtas parametrų parinkimo bandymas

A screenshot of a computer

Description automatically generated

**7 pav.** Penktas parametrų parinkimo bandymas

Geram neuroninio tinkle mokymui, viena iš užduočių buvo parinkti atitinkamus parametrus su kuriai būtų rastas geriausias sprendimas. Komponentėje "Multilayer Perceptron" buvo keičiamas paslėptų neuronų skaičius (hiddenLayers), mokymo greičio parametras (learningRate) ir inercijos koeficientas (momentum). Pradinė paslėptų sluoksnių reikšmė buvo nustatyta kaip "a", reiškianti vieną automatiškai parinktą sluoksnį. Norint sukurti tinklą su keliais paslėptais sluoksniais, kiekvieno sluoksnio neuronų skaičius nurodomas atskiriant kableliais, pavyzdžiui, formatu „2,4“, reiškiančiu du sluoksnius su atitinkamai 2 ir 4 neuroniais kiekviename.

Pagal viršuje pateiktus bandymus, buvo rastas geriausia sprendimas. Mokymo greitis geriausius sprendimus davė arti 0,3, inercijos koeficientas arti 0,15. Paslėptų neuronų skaičius buvo parinktas 4 ir 4, tai yra 2 paslėptų sluoksnių neuroninis tinklas. Su šiais parametrais modelis galėjo pasiekti 115 teisingų klasifikuotų duomenų iš 120 (95,833%).

A screenshot of a computer

Description automatically generated

**8 pav.** Pirmos sekos neuroninio tinklo vaizdas

Kaip pateikta **8 pav.** Pirmos sekos neuroninio tinklo vaizdas galima matyti 2 raudona spalva pažymėtus paslėptus neuroninio tinklos sluoksnius.

A screenshot of a computer

Description automatically generated

**9 pav.** Duomenų požymių porų vaizda Dekarto koordinačių sistemoje

**1 pav.** iris\_train\_test.arff failo pavyzdys**2 pav.** Pirma užduočių seka esančioje sekoje galima pastebėti ScatterPlotMatrix komponentą, kuris atvaizduoja duomenų požymių poras koordinačių sistemoje. Tai leidžia aiškiau matyti sugrupuotus duomenis ir jų tarpusavio ryšį. Šias duomenų poras ir galima matyti **9 pav.** Duomenų požymių porų vaizda Dekarto koordinačių sistemoje

# Antra užduočių seka

A diagram of a diagram

Description automatically generated

**10 pav.** Antra užduočių seka

Antra užduoties seka, pavaizduota **10 pav.** Antra užduočių seka, buvo sukurta patikrinti pirmoje užduočių sekoje išmokytą klasifikavimo modelį. Panašiai kaip ir pirmoje sekoje buvo panaudoti tie patys komponentai, bet papildomai įtraukti šie komponentai: TestSetMaker ir PredictionAppender. Į ArffLoader komponentą įkeltas *iris\_new.arff* duomenų rinkinys. Paleidus šią seką gaunami rezultatai – klasifikavimo duomenys ir klasifikavimo tikslumo metrikos.

A screenshot of a computer

Description automatically generated

**11 pav.** Modelio klasifikavimo duomenys

A screenshot of a computer program

Description automatically generated

**12 pav.** Klasifikavimo tikslumo metrikos

**11 pav.** Modelio klasifikavimo duomenysparodyti klasifikavimo duomenys parodo modelio prognozes kiekvienai klasei šalia duotų duomenų, kurie buvo pateikti iš *iris\_new.arff* failo. Kad patikrinti jų tikslumą ar teisingai buvo suklasifikuoti duomenys galima patikrinti klasifikavimo tikslumo metrikas, kurios yra pateiktos **12 pav.** Klasifikavimo tikslumo metrikos. Iš pateiktų metrikų galima pastebėti, kad visi testavimo duomenys buvo suklasifikuoti teisingai ir modelis buvo gerai išmokytas.

# Trečia užduočių seka

A diagram of a network

Description automatically generated

**13 pav.** Trečia užduočių seka

Ši užduočių seka sukurta WEKA aplinkoje skiriasi nuo kitų tuo, kad užduotis sukurta su dviem ArffLoader komponentais: į vieną buvo įkeltas *iris\_new.arff* duomenų rinkinys, o į kitą – *iris\_train\_test.arff*. Paleidus šią seką vienu metu yra mokomas ir testuojamas dirbtinio neuroninio tinklo modelis. Gauti rezultatai yra pateikiami kaip klasifikavimo tikslumo metrikos ir gauti svoriai išmokius modelį. Kaip pavaizduota **15 pav.** Po apmokymo gauti parametrų svoriai, po apmokymo gauti parametrų svoriai yra pateikiami kiekvieno dirbtinio neuroninio tinklo sluoksnio neuronui. Naudojant šiuos svorius ir sukonstravus matematinį modelį su tais pačiais sluoksniais galima atkartoti tuos pačius rezultatus, pateiktus **14 pav.** Nauji modelio klasifikavimo duomenys, panaudojant *iris\_new.arff* faile esančius duomenis.

A screenshot of a computer

Description automatically generated

**14 pav.** Nauji modelio klasifikavimo duomenys

A close-up of a computer code

Description automatically generated

**15 pav.** Po apmokymo gauti parametrų svoriai

# MS Excel dirbtinio neuroninio tinklo modelis

Kaip paminėta ankstesniame skyriuje, kad panaudojant sukurtus svorius galima atkartoti tą patį ar panašų rezultatą. Tai patikrinti buvo panaudota MS Excel programa. Pirmasis žingsnis buvo normalizuoti įvesties duomenis iš *iris\_new.arff* failo.

**1 lentelė.** Įvesties duomenys ir jų normavimas

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **6.1 Duomenys** | |  |  |  | **6.2 Normavimas** | |  |
| Sepal length | Petal length | Petal width | Class |  | Sepal length | Petal length | Petal width |
| 5 | 1,3 | 0,3 | Iris-setosa |  | -0,52 | -1 | -0,91304 |
| 4,5 | 1,3 | 0,3 | Iris-setosa |  | -0,92 | -1 | -0,91304 |
| 4,4 | 1,3 | 0,2 | Iris-setosa |  | -1 | -1 | -1 |
| 5 | 1,6 | 0,6 | Iris-setosa |  | -0,52 | -0,86957 | -0,65217 |
| 5,1 | 1,9 | 0,4 | Iris-setosa |  | -0,44 | -0,73913 | -0,82609 |
| 4,8 | 1,4 | 0,3 | Iris-setosa |  | -0,68 | -0,95652 | -0,91304 |
| 5,1 | 1,6 | 0,2 | Iris-setosa |  | -0,44 | -0,86957 | -1 |
| 4,6 | 1,4 | 0,2 | Iris-setosa |  | -0,84 | -0,95652 | -1 |
| 5,3 | 1,5 | 0,2 | Iris-setosa |  | -0,28 | -0,91304 | -1 |
| 5 | 1,4 | 0,2 | Iris-setosa |  | -0,52 | -0,95652 | -1 |
| 5,5 | 4,4 | 1,2 | Iris-versicolor |  | -0,12 | 0,347826 | -0,13043 |
| 6,1 | 4,6 | 1,4 | Iris-versicolor |  | 0,36 | 0,434783 | 0,043478 |
| 5,8 | 4 | 1,2 | Iris-versicolor |  | 0,12 | 0,173913 | -0,13043 |
| 5 | 3,3 | 1 | Iris-versicolor |  | -0,52 | -0,13043 | -0,30435 |
| 5,6 | 4,2 | 1,3 | Iris-versicolor |  | -0,04 | 0,26087 | -0,04348 |
| 5,7 | 4,2 | 1,2 | Iris-versicolor |  | 0,04 | 0,26087 | -0,13043 |
| 5,7 | 4,2 | 1,3 | Iris-versicolor |  | 0,04 | 0,26087 | -0,04348 |
| 6,2 | 4,3 | 1,3 | Iris-versicolor |  | 0,44 | 0,304348 | -0,04348 |
| 5,1 | 3 | 1,1 | Iris-versicolor |  | -0,44 | -0,26087 | -0,21739 |
| 5,7 | 4,1 | 1,3 | Iris-versicolor |  | 0,04 | 0,217391 | -0,04348 |
| 6,7 | 5,6 | 2,4 | Iris-virginica |  | 0,84 | 0,869565 | 0,913043 |
| 6,9 | 5,1 | 2,3 | Iris-virginica |  | 1 | 0,652174 | 0,826087 |
| 5,8 | 5,1 | 1,9 | Iris-virginica |  | 0,12 | 0,652174 | 0,478261 |
| 6,8 | 5,9 | 2,3 | Iris-virginica |  | 0,92 | 1 | 0,826087 |
| 6,7 | 5,7 | 2,5 | Iris-virginica |  | 0,84 | 0,913043 | 1 |
| 6,7 | 5,2 | 2,3 | Iris-virginica |  | 0,84 | 0,695652 | 0,826087 |
| 6,3 | 5 | 1,9 | Iris-virginica |  | 0,52 | 0,608696 | 0,478261 |
| 6,5 | 5,2 | 2 | Iris-virginica |  | 0,68 | 0,695652 | 0,565217 |
| 6,2 | 5,4 | 2,3 | Iris-virginica |  | 0,44 | 0,782609 | 0,826087 |
| 5,9 | 5,1 | 1,8 | Iris-virginica |  | 0,2 | 0,652174 | 0,391304 |

Normavimas buvo atliktas naudojant formulę **16 pav.** Normavimo formulė.

A black and white math equation

Description automatically generated with medium confidence

**16 pav.** Normavimo formulė

**2 lentelė.** Neuronų jungčių tarp įvesties ir paslėpto sluoksnio neuronų svoriai

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Jungčių tarp įvesties ir paslėpto sluoksnio neuronų** | | | | |
| Neuron | Threshhold | Sepal length | Petal length | Petal width |
| 1 | -4,28022516444049 | -3,09397289167281 | 11,51554304708630 | 4,41533596253539 |
| 2 | 3,07274912441607 | 2,37191962164487 | -8,50570848393368 | -3,20124134488006 |
| 3 | -3,77578062704813 | -0,19227544736188 | -4,66993077323038 | -4,16439146783643 |
| 4 | 1,69607240290226 | 1,41644080160431 | -4,75294435130322 | -1,82099635824253 |
| 5 | 2,22809170703311 | -0,37426367113162 | 3,18749357863442 | 2,79902639944813 |

**3 lentelė.** Neuronų jungčių tarp paslėpto sluoksnio ir išėjimo neuronų svoriai

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Jungčių tarp paslėpto sluoksnio neuronų ir išėjimų** | | | | | | |
| Neuron | Threshold | w1 | w2 | w3 | w4 | w5 |
| 1 | -1,69700419071657 | -2,61271965505448 | 0,81718583158522 | 5,33311969769728 | 0,69011618453227 | -5,26174700954742 |
| 2 | -0,83150533364569 | -8,21579694678083 | 4,35441508109297 | -9,51101811614322 | 1,24907558517973 | 3,63369189133528 |
| 3 | -1,59030537656810 | 6,35989451393923 | -6,30635467163820 | -3,08503591554279 | -4,34284068483260 | 2,18370166653828 |

Duomenys pateikti **1 lentelė.** Įvesties duomenys ir jų normavimas ir **2 lentelė.** Neuronų jungčių tarp įvesties ir paslėpto sluoksnio neuronų svoriai lentelėse atvaizduoja gautų neuronų svorių reikšmes, kad būtų aišku, kurio sluoksnio kuris svorių rinkinys yra. Šie duomenys buvo paimti iš **15 pav.** Po apmokymo gauti parametrų svoriai esančių duomenų. Turint visus pateiktus svorius ir įvesties duomenis jau galima atlikti atitinkamus skaičiavimus kaip svorių ir įvesties duomenų sandaugų sumų skaičiavimas ir sigmoidinių funkcijų pritaikymas gautiems rezultatams. Šiuos skaičiavimus galima pasižiūrėti pateiktame MS Excel faile. Gauti rezultatai yra gautos DNT prognozės kiekvienai klasei visiems įvesties duomenims.

**4 lentelė.** Excel DNT ir WEKA DNT gauti rezultatai

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **6.7 Neuroninio tinklo išėjimai** | | |  | **WEKA data** | | |
| Iris-setosa | Iris-versicolor | Iris-virginica |  | Iris-setosa | Iris-versicolor | Iris-virginica |
| 0,992723 | 0,010611 | 0,000000 |  | 0,987311 | 0,012689 | 0 |
| 0,992536 | 0,010751 | 0,000000 |  | 0,987146 | 0,012854 | 0 |
| 0,992912 | 0,010235 | 0,000000 |  | 0,988255 | 0,011745 | 0 |
| 0,987573 | 0,018688 | 0,000000 |  | 0,969918 | 0,030082 | 0 |
| 0,988374 | 0,017443 | 0,000000 |  | 0,975204 | 0,024796 | 0 |
| 0,992376 | 0,011047 | 0,000000 |  | 0,98658 | 0,013419 | 0 |
| 0,992428 | 0,011055 | 0,000000 |  | 0,986813 | 0,013187 | 0 |
| 0,992764 | 0,010465 | 0,000000 |  | 0,987845 | 0,012155 | 0 |
| 0,992764 | 0,010618 | 0,000000 |  | 0,987504 | 0,012496 | 0 |
| 0,992895 | 0,010366 | 0,000000 |  | 0,98796 | 0,01204 | 0 |
| 0,000861 | 0,802371 | 0,211991 |  | 0,003106 | 0,995408 | 0,001487 |
| 0,000606 | 0,690854 | 0,364698 |  | 0,001975 | 0,993757 | 0,004267 |
| 0,004730 | 0,998272 | 0,000467 |  | 0,00704 | 0,992852 | 0,000108 |
| 0,027596 | 0,996927 | 0,000038 |  | 0,035577 | 0,9644 | 0,000023 |
| 0,001666 | 0,973523 | 0,022207 |  | 0,003909 | 0,995547 | 0,000544 |
| 0,002622 | 0,991859 | 0,004631 |  | 0,005158 | 0,994583 | 0,000259 |
| 0,001959 | 0,984468 | 0,011655 |  | 0,004153 | 0,995423 | 0,000424 |
| 0,002836 | 0,996068 | 0,002087 |  | 0,004621 | 0,995142 | 0,000237 |
| 0,041920 | 0,996234 | 0,000021 |  | 0,044234 | 0,955749 | 0,000017 |
| 0,002664 | 0,993620 | 0,003636 |  | 0,00484 | 0,9949 | 0,00026 |
| 0,000073 | 0,004784 | 0,998784 |  | 0,000078 | 0,005733 | 0,994189 |
| 0,000086 | 0,007280 | 0,997198 |  | 0,000114 | 0,019695 | 0,98019 |
| 0,000079 | 0,005485 | 0,998303 |  | 0,000103 | 0,012752 | 0,987145 |
| 0,000072 | 0,004665 | 0,998868 |  | 0,000075 | 0,005287 | 0,994638 |
| 0,000072 | 0,004664 | 0,998872 |  | 0,000075 | 0,005224 | 0,994701 |
| 0,000079 | 0,005781 | 0,998128 |  | 0,000096 | 0,010874 | 0,98903 |
| 0,000094 | 0,008756 | 0,996330 |  | 0,000161 | 0,055187 | 0,944652 |
| 0,000083 | 0,006511 | 0,997669 |  | 0,000116 | 0,019548 | 0,980336 |
| 0,000073 | 0,004787 | 0,998780 |  | 0,000079 | 0,005969 | 0,993952 |
| 0,000082 | 0,006071 | 0,997932 |  | 0,000121 | 0,020639 | 0,97924 |

Pagal gautus duomenis galima pastebėti, kad gauti rezultatai Excel faile yra artimi gautiems WEKA sistemoje ištreniruotam modeliui. Tai įrodo, kad naudojant tuos pačius svorius ir neuroninio tinklo modelio architektūrą galima atkartoti tuos pačius ar panašius rezultatus.