

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFOMATIKOS FAKULTETAS
DUOMENŲ MOKSLO BAKALAU'RAS

Tiesioginio sklidimo DNT naudojant sistemą WEKA

Ataskaita

Atliko: Simona Gelžinytė 3 k. 2 gr.

Vadovas: dr. Viktor Medvedev

Vilnius, 2022

TURINYS

ĮVADAS.....	3
Tyrimo tikslas	3
Tyrimo uždaviniai	3
Duomenys ir naudota programinė įranga.....	3
PAGRINDINĖ DALIS	4
Duomenų paruošimas.....	4
Užduočių sekos sukonstravimas	5
Neuroninio tinklo parametru parinkimas	7
Naujų duomenų klasifikavimas.....	11
Duomenų požymių porų vaizdai Dekarto koordinačių sistemoje.....	13
Klasifikavimas ir testavimas	14
Neuronų išėjimo reikšmių perskaičiavimas MS Excel programoje.....	17
Duomenų nukopijavimas į MS Excel programą ir normavimas.....	17
Neuronų svorių lentelės	18
Duomenų jėjimo vektorių ir paslėptų neuronų svorių vektorių sandaugų sumos.....	18
Sigmoidinės funkcijos reikšmių skaičiavimas nuo gautų sumų	20
Sigmoidinės funkcijos gautų reikšmių (t.y. paslėpto sluoksnio išėjimų) vektorių ir paslėptų neuronų svorių vektorių sandaugų sumų skaičiavimas.....	22
Tikimybės	23
Gautų tikimybių palyginimas su neuroninio tinklo gautomis tikimybėmis.....	24
IŠVADOS	25

ĮVADAS

Tyrimo tikslas

Išmokyti neuroninį tinklą teisingai klasifikuoti duomenis naudojant sistemą WEKA.

Tyrimo uždaviniai

- Paruošti duomenis
- Sudaryti užduočių sekas, naudojant WEKA programą
- Klasifikuoti naujus duomenis
- Gautus rezultatus, naudojant WEKA programa, apskaičiuoti taip pat su MS Excel programa
- Palyginti WEKOS ir MS Excel gautus rezultatus

Duomenys ir naudota programinė įranga

Šiame tyrime buvo naudoti duomenys iš internetinės svetainės „UCI Machine Learning Repository“. Buvo pasirinkta duomenys apie irisus (<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/>).

Padalinti duomenims į du failus buvo naudojama funkcija train_test_split. Iš pradžių kiekviena rūšis padalinama santykiu 80:20. Kai kiekviena rūšis padalijama į dvi aibes, tuomet visų rūsių failai yra sujungiami naudojant funkciją pd.concat.

Šioje praktinėje užduotyje buvo naudojama duomenų tyrybos sistema WEKA ir „Python“ programavimo kalba, duomenims padalinti į du failus.

PAGRINDINĖ DALIS

Duomenų paruošimas

```
from scipy.io import arff as sciarff
import pandas as pd
from sklearn.model_selection import train_test_split
from collections import Counter

data = sciarff.loadarff('/Users/simonagelzinyte/Documents/Duomenų mokslas/5
semestras/AI/iris.arff')
df = pd.DataFrame(data[0])

setosa = df[df['class'].isin([b'Iris-setosa'])]
versicolor = df[df['class'].isin([b'Iris-versicolor'])]
virginica = df[df['class'].isin([b'Iris-virginica'])]

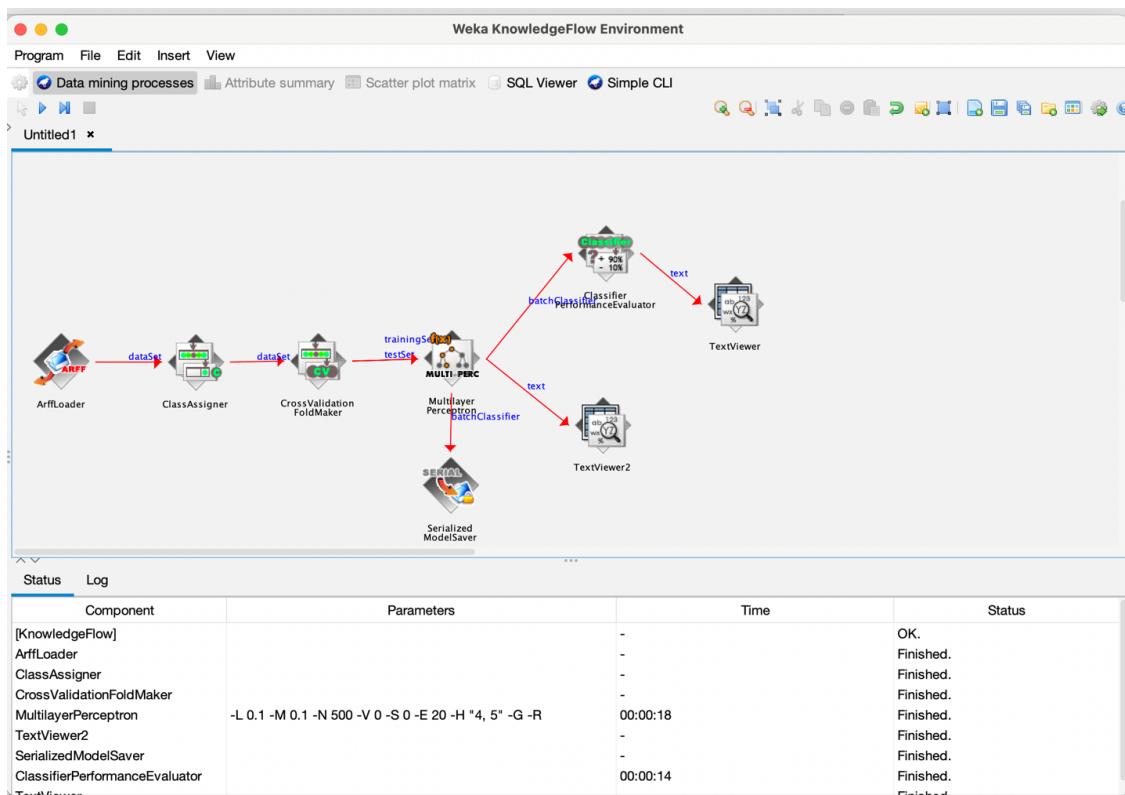
train_se, test_se = train_test_split(setosa, test_size= 0.2, random_state=
3)
train_ve, test_ve = train_test_split(versicolor, test_size= 0.2,
random_state= 3)
train_vi, test_vi = train_test_split(virginica, test_size= 0.2,
random_state= 3)

train = pd.concat([train_se, train_ve, train_vi])
test = pd.concat([test_se, test_ve, test_vi])

from pandas2arff import pandas2arff
pandas2arff(train, 'iris_train_test.arff', wekaname = "iris_train_test")
pandas2arff(test, 'iris_new.arff', wekaname = "iris_new")
```

Visą duomenų aibę sudaro 150 įrašų (eilučių). Mokymo ir testavimo aibės santykis 80:20, t.y. mokymo aibė (*iris_train_test.arff*) sudaryta iš 120 įrašų, o testavimo (*iris_new.arff*) sudaryta iš 30.

Užduočių sekos sukonstravimas



I pav pirmos užduočių sekos sukonstravimas

ArffLoarder – leidžia įkelti duomenis arff formatu.

ClassAssigner – nurodome, kuris duomenų požymis yra klasė.

CrossValidation FoldMaker – atliekama kryžminė patikra.

Multilayer Perceptron – bus naudojamas daugiasluoksnis perceptronas.

Classifier PerformanceEvaluator – gauname klasifikavimo rezultatų tikslumą.

Serialized ModelSaver – išsaugome modelį.

TextViewer – gautų duomenų pateikimas.

Buvo pasirinkta:

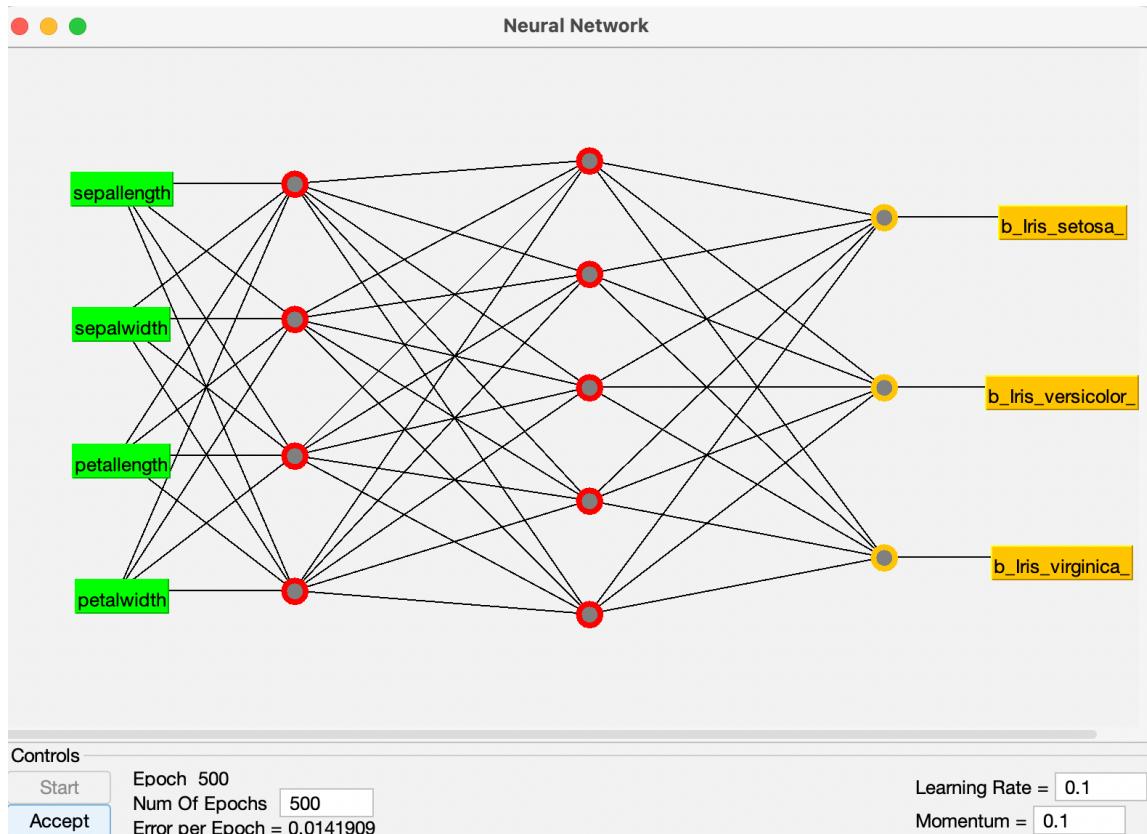
Layers – 4;5

LearningRate – 0,1

Momentum – 0,1

Accuracy – 96,6667%

Neuroninis tinklas:



2 pav neuroninis tinklas.

Rezultatų išvedimas į ekraną:

```
Text
==== Evaluation result ===

Scheme: MultilayerPerceptron
Options: -L 0.1 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H "4, 5" -G -R
Relation: iris_train_test

==== Summary ===

Correctly Classified Instances      116          96.6667 %
Incorrectly Classified Instances    4           3.3333 %
Kappa statistic                   0.95
Mean absolute error               0.055
Root mean squared error           0.1397
Relative absolute error            12.3767 %
Root relative squared error      29.6399 %
Total Number of Instances         120

==== Detailed Accuracy By Class ===

      TP Rate   FP Rate   Precision   Recall   F-Measure   MCC   ROC Area   PRC Area   Class
        1.000     0.000     1.000     1.000     1.000     1.000     1.000     1.000   b_Iris_setosa_
        0.925     0.013     0.974     0.925     0.949     0.925     0.993     0.988   b_Iris_versicolor_
        0.975     0.038     0.929     0.975     0.951     0.927     0.993     0.987   b_Iris_virginica_
Weighted Avg.                     0.967     0.017     0.967     0.967     0.967     0.950     0.996     0.992

==== Confusion Matrix ===

 a   b   c   <-- classified as
40   0   0   |   a = b_Iris_setosa_
 0  37   3   |   b = b_Iris_versicolor_
 0   1  39   |   c = b_Iris_virginica_
```

3 pav. tiksliausi klasifikavimo rezultatai

Neuroninio tinklo parametrų parinkimas

1) Layers – 4;5

LearningRate – 0,5

Momentum – 0,1

Accuracy – 93,3333%

Rezultatų išvedimas į ekraną:

```

Text
==== Evaluation result ====
Scheme: MultilayerPerceptron
Options: -L 0.5 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H "4, 5" -G -R
Relation: iris_train_test

==== Summary ====
Correctly Classified Instances      112          93.3333 %
Incorrectly Classified Instances    8           6.6667 %
Kappa statistic                   0.9
Mean absolute error               0.0501
Root mean squared error          0.1917
Relative absolute error           11.2613 %
Root relative squared error     40.6758 %
Total Number of Instances        120

==== Detailed Accuracy By Class ====
      TP Rate   FP Rate   Precision   Recall   F-Measure   MCC   ROC Area   PRC Area   Class
  1.000    0.000    1.000     1.000    1.000    1.000    1.000    1.000   b_Iris_setosa_
  0.875    0.038    0.921     0.875    0.897    0.849    0.953    0.940   b_Iris_versicolor_
  0.925    0.063    0.881     0.925    0.902    0.852    0.985    0.969   b_Iris_virginica_
Weighted Avg.    0.933    0.033    0.934     0.933    0.933    0.900    0.979    0.970

==== Confusion Matrix ====
  a   b   c   <-- classified as
40   0   0   |   a = b_Iris_setosa_
  0  35   5   |   b = b_Iris_versicolor_
  0   3  37   |   c = b_Iris_virginica_

```

4 pav. parametrų parinkimas, 1 bandymas

2) Layers – 4;5

LearningRate – 0,4

Momentum – 0,4

Accuracy – 92,5%

Rezultatų išvedimas į ekrana:

```

Text
==== Evaluation result ===

Scheme: MultilayerPerceptron
Options: -L 0.4 -M 0.4 -N 500 -V 0 -S 0 -E 20 -H "4, 5" -G -R
Relation: iris_train_test

==== Summary ===

Correctly Classified Instances      111          92.5    %
Incorrectly Classified Instances     9           7.5    %
Kappa statistic                   0.8875
Mean absolute error                0.0527
Root mean squared error            0.1976
Relative absolute error             11.8613 %
Root relative squared error        41.9143 %
Total Number of Instances          120

==== Detailed Accuracy By Class ===

      TP Rate   FP Rate   Precision   Recall   F-Measure   MCC   ROC Area   PRC Area   Class
      1.000     0.000     1.000      1.000     1.000     1.000   1.000     1.000   b_Iris_setosa_
      0.875     0.050     0.897      0.875     0.886     0.830   0.955     0.946   b_Iris_versicolor_
      0.900     0.063     0.878      0.900     0.889     0.832   0.985     0.968   b_Iris_virginica_
Weighted Avg.      0.925     0.038     0.925      0.925     0.925     0.888   0.980     0.971

==== Confusion Matrix ===

  a   b   c   <-- classified as
40   0   0   |   a = b_Iris_setosa_
  0  35   5   |   b = b_Iris_versicolor_
  0   4  36   |   c = b_Iris_virginica_

```

5 pav. parametry parinkimas, 2 bandymas

3) Layers - 4

LearningRate – 0,4

Momentum – 0,4

Accuracy – 95,8333%

Rezultatų išvedimas į ekraną:

Text

```
==== Evaluation result ===  
Scheme: MultilayerPerceptron  
Options: -L 0.4 -M 0.4 -N 500 -V 0 -S 0 -E 20 -H 4 -G -R  
Relation: iris_train_test  
==== Summary ===  
Correctly Classified Instances      115          95.8333 %  
Incorrectly Classified Instances     5           4.1667 %  
Kappa statistic                   0.9375  
Mean absolute error                0.043  
Root mean squared error            0.1705  
Relative absolute error             9.6689 %  
Root relative squared error        36.178 %  
Total Number of Instances          120  
==== Detailed Accuracy By Class ===  
          TP Rate   FP Rate   Precision   Recall   F-Measure   MCC    ROC Area   PRC Area   Class  
          1.000     0.000     1.000      1.000     1.000      1.000   1.000     1.000   b_Iris_setosa_  
          0.925     0.025     0.949      0.925     0.937      0.906   0.965     0.949   b_Iris_versicolor_  
          0.950     0.038     0.927      0.950     0.938      0.907   0.984     0.964   b_Iris_virginica_  
Weighted Avg.      0.958     0.021     0.959      0.958     0.958      0.938   0.983     0.971  
==== Confusion Matrix ====  
    a   b   c   <-- classified as  
40  0  0 |  a = b_Iris_setosa_  
 0 37  3 |  b = b_Iris_versicolor_  
 0  2 38 |  c = b_Iris_virginica_
```

6 pav. parametrų parinkimas, 3 bandymas

Iš gautų rezultatų, matome, jog tiksliausi klasifikavimo rezultatai gaunami, kai:

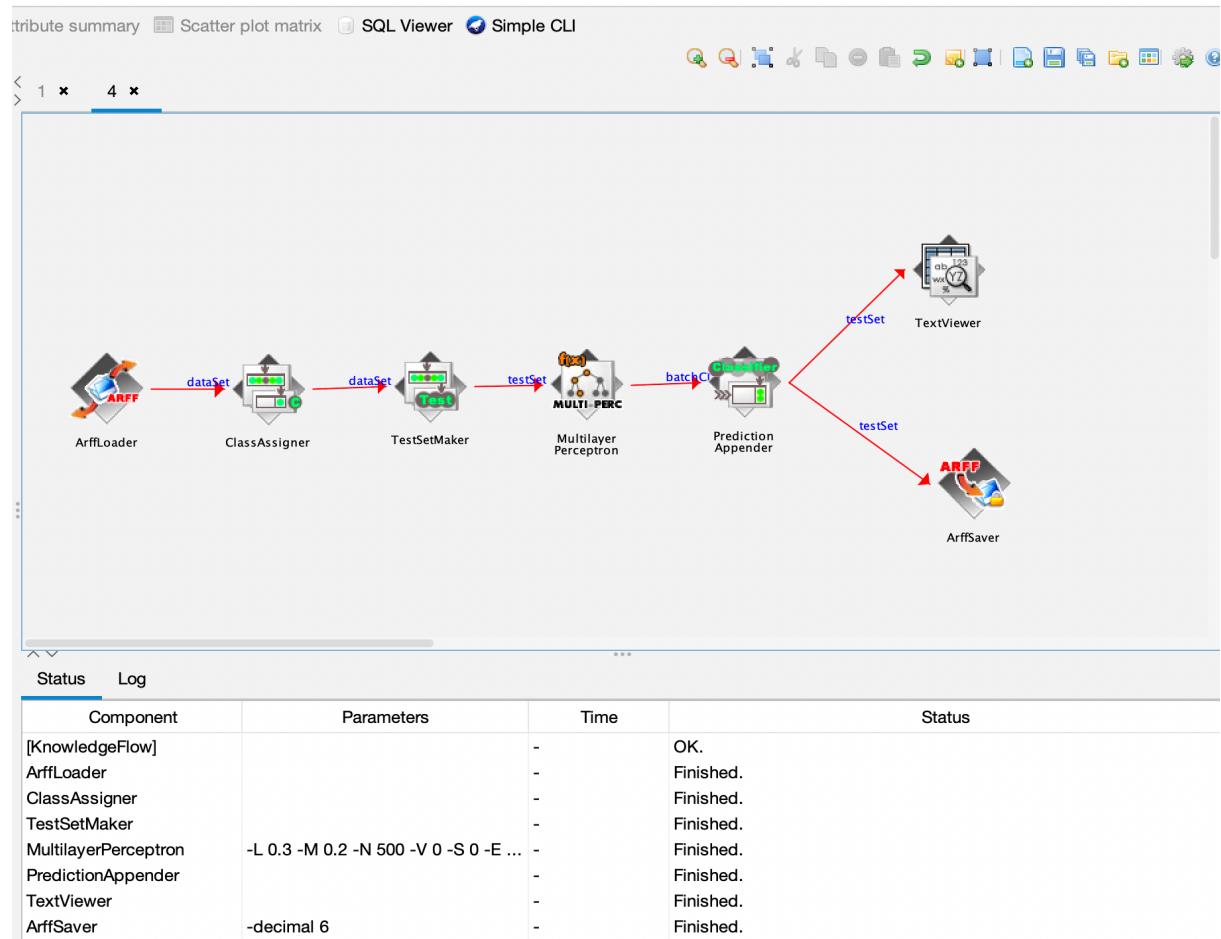
Layers – 4;5

LearningRate – 0,1

Momentum – 0,1

Accuracy – 96,6667%

Naujų duomenų klasifikavimas



7 pav. naujų duomenų klasifikavimo užduočių seka

Prediction Appender – prognozuoja kokiai klasei bus priskirta.

Text

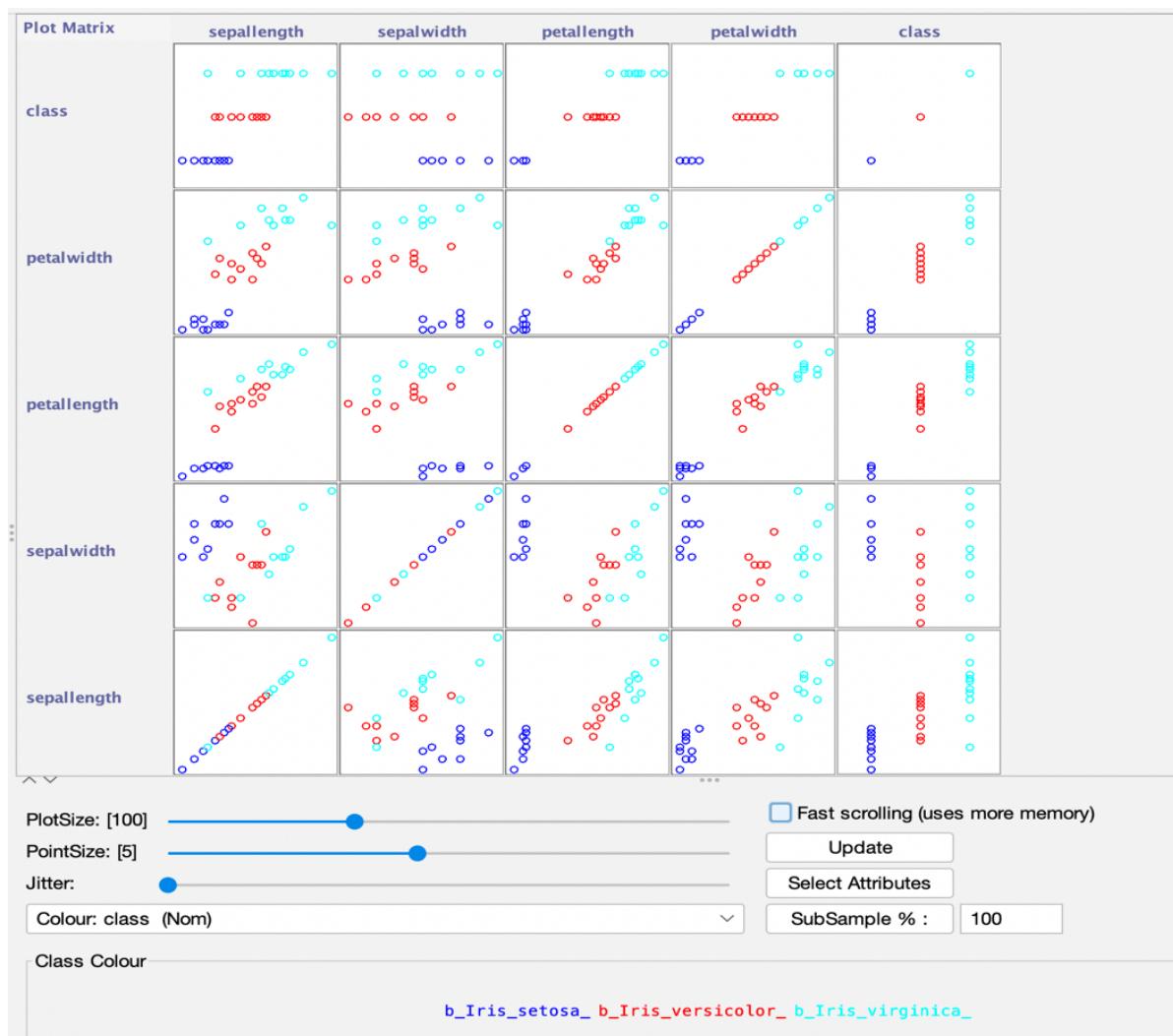
```
@relation iris_new_set_1_of_1

@attribute sepallength numeric
@attribute sepalwidth numeric
@attribute petallength numeric
@attribute petalwidth numeric
@attribute class {b_Iris_setosa_,b_Iris_versicolor_,b_Iris_virginica_}
@attribute 'class_predicted_by: MultilayerPerceptron' {b_Iris_setosa_,b_Iris_versicolor_,b_Iris_virginica_}

@data
4.8,3,1.4,0.1,b_Iris_setosa_,b_Iris_setosa_
5.1,3.4,1.5,0.2,b_Iris_setosa_,b_Iris_setosa_
4.9,3.1,1.5,0.1,b_Iris_setosa_,b_Iris_setosa_
4.6,3.2,1.4,0.2,b_Iris_setosa_,b_Iris_setosa_
5.4,3.4,1.5,0.4,b_Iris_setosa_,b_Iris_setosa_
5.2,3.4,1.4,0.2,b_Iris_setosa_,b_Iris_setosa_
4.3,3,1.1,0.1,b_Iris_setosa_,b_Iris_setosa_
5.3,3.7,1.5,0.2,b_Iris_setosa_,b_Iris_setosa_
4.8,3,1.4,0.3,b_Iris_setosa_,b_Iris_setosa_
4.6,3.4,1.4,0.3,b_Iris_setosa_,b_Iris_setosa_
6.2,2,4,1,b_Iris_versicolor_,b_Iris_versicolor_
5.5,2.5,4,1.3,b_Iris_versicolor_,b_Iris_versicolor_
5.2,2.7,3.9,1.4,b_Iris_versicolor_,b_Iris_versicolor_
6.2,2.9,4.3,1.3,b_Iris_versicolor_,b_Iris_versicolor_
5.5,2.4,3.7,1,b_Iris_versicolor_,b_Iris_versicolor_
6.2,9,4.5,1.5,b_Iris_versicolor_,b_Iris_versicolor_
6.1,2.9,4.7,1.4,b_Iris_versicolor_,b_Iris_versicolor_
5.1,2.5,3,1.1,b_Iris_versicolor_,b_Iris_versicolor_
5.7,3,4.2,1.2,b_Iris_versicolor_,b_Iris_versicolor_
6.3,3.3,4.7,1.6,b_Iris_versicolor_,b_Iris_versicolor_
6.8,3,5.5,2.1,b_Iris_virginica_,b_Iris_virginica_
6.9,3.1,5.4,2.1,b_Iris_virginica_,b_Iris_virginica_
7.2,3.6,6.1,2.5,b_Iris_virginica_,b_Iris_virginica_
6.5,3.5,2,2,b_Iris_virginica_,b_Iris_virginica_
7.9,3.8,6.4,2,b_Iris_virginica_,b_Iris_virginica_
6.4,2.8,5.6,2.1,b_Iris_virginica_,b_Iris_virginica_
5.7,2.5,5,2,b_Iris_virginica_,b_Iris_virginica_
6.2,3.4,5.4,2.3,b_Iris_virginica_,b_Iris_virginica_
6.7,3.5,2,2.3,b_Iris_virginica_,b_Iris_virginica_
4.9,2.5,4.5,1.7,b_Iris_virginica_,b_Iris_virginica_
```

8 pav. rezultatai su prognozuojamu klasių priskyrimu

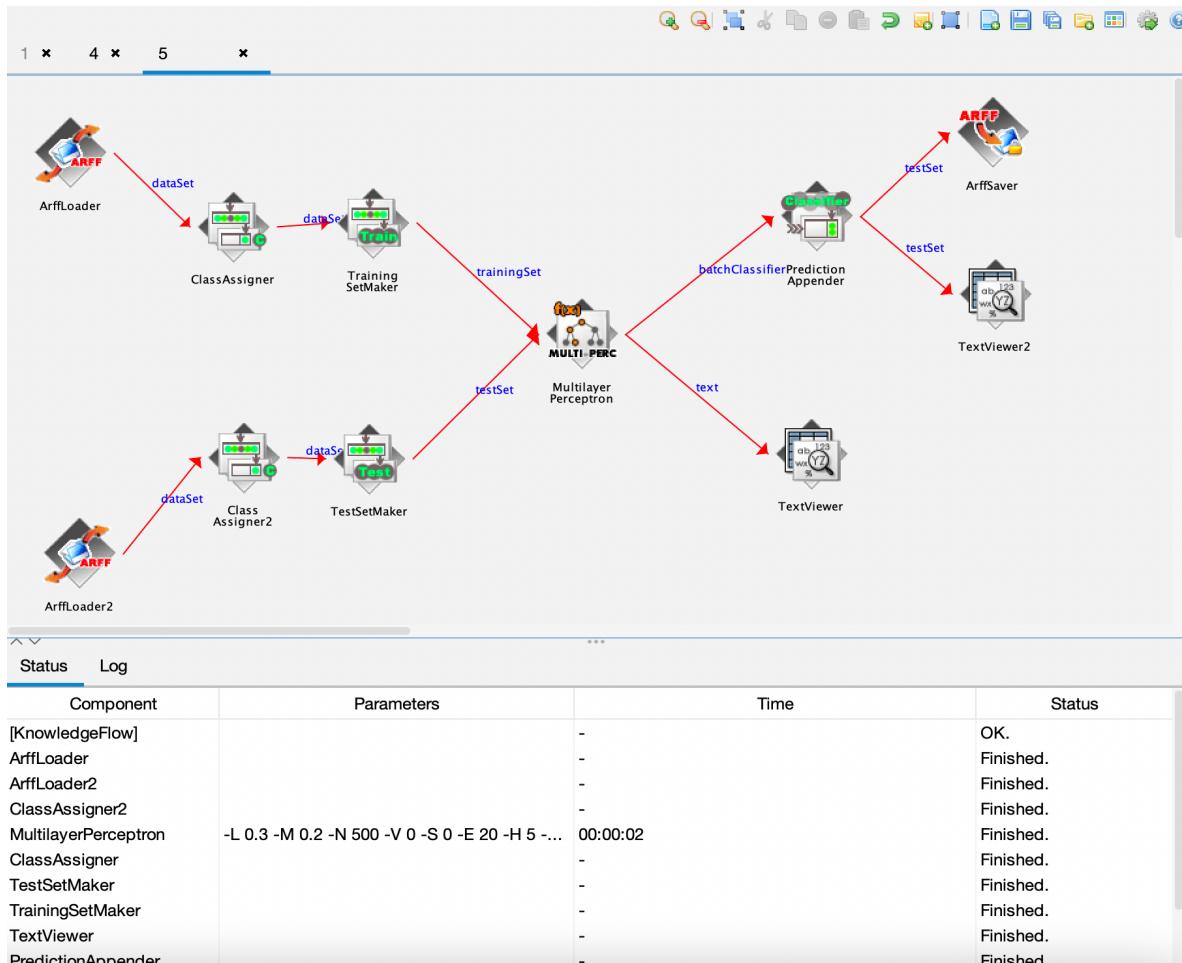
Duomenų požymių porų vaizdai Dekarto koordinačių sistemoje



9 pav. Dekarto koordinačių sistema

Klasifikavimas ir testavimas

Paslėptų neuronų skaičius: 5



10 pav. duomenų klasifikavimas ir testavimas, užduočių seka

Training SetMaker – duomenys priskiriami mokymo aibei.

Test SetMaker – duomenys priskiriami testavimo aibei.

Gauti svoriai:

==== Classifier model ====

Scheme: MultilayerPerceptron

Relation: iris_train_test

Sigmoid Node 0

Inputs Weights

Threshold 0.33979280896170877

Node 3 4.877102706276549

Node 4 -2.0338118954920206
Node 5 -1.6409348811996538
Node 6 -1.65852242064022
Node 7 -4.7040621352627685

Sigmoid Node 1

Inputs Weights
Threshold 0.13348262568013625
Node 3 -5.188271400974106
Node 4 0.45600560151121816
Node 5 -6.1621726780645245
Node 6 -6.456806138828598
Node 7 5.521804603919687

Sigmoid Node 2

Inputs Weights
Threshold -5.340176446560711
Node 3 -5.196332019273902
Node 4 0.4211044030122157
Node 5 5.8232945615987495
Node 6 6.001651174481241
Node 7 -0.3152024533057182

Sigmoid Node 3

Inputs Weights
Threshold -1.686351728719476
Attrib sepallength -0.08435646995823146
Attrib sepalwidth 2.0970469171540143
Attrib petallength -2.869447922100432
Attrib petalwidth -2.933471422299198

Sigmoid Node 4

Inputs Weights
Threshold -0.2829199207803021
Attrib sepallength 0.3378929326643106
Attrib sepalwidth -0.4356287470673166
Attrib petallength 0.9598467239015573
Attrib petalwidth 1.0947936468429507

Sigmoid Node 5

Inputs Weights
Threshold -5.2338533062720405
Attrib sepallength -0.9088499963731073
Attrib sepalwidth -2.167358227119635
Attrib petallength 7.308737766181561
Attrib petalwidth 6.166506686221638

Sigmoid Node 6

Inputs Weights

```

Threshold -5.353932991833778
Attrib sepallength -1.0008614530094377
Attrib sepalwidth -2.284039115131845
Attrib petallength 7.647405361642313
Attrib petalwidth 6.294658034934279

```

Sigmoid Node 7

Inputs Weights

```

Threshold 2.643956440305184
Attrib sepallength 0.9326692959120354
Attrib sepalwidth -2.0813381931714523
Attrib petallength 3.1228326512327915
Attrib petalwidth 3.128614534687554

```

Class b_Iris_setosa_

Input

Node 0

Class b_Iris_versicolor_

Input

Node 1

Class b_Iris_virginica_

Input

Node 2

Gautos tikimybės:

```

@relation iris_new_set_1_of_1

@attribute sepallength numeric
@attribute sepalwidth numeric
@attribute petallength numeric
@attribute petalwidth numeric
@attribute class {b_Iris_setosa_,b_Iris_versicolor_,b_Iris_virginica_}
@attribute MultilayerPerceptron_prob_b_Iris_setosa_ numeric
@attribute MultilayerPerceptron_prob_b_Iris_versicolor_ numeric
@attribute MultilayerPerceptron_prob_b_Iris_virginica_ numeric

@data
4.8,3,1.4,0.1,b_Iris_setosa_,0.990964,0.009004,0.000032
5.1,3.4,1.5,0.2,b_Iris_setosa_,0.991537,0.008433,0.000031
4.9,3.1,1.5,0.1,b_Iris_setosa_,0.991071,0.008898,0.000032
4.6,3.2,1.4,0.2,b_Iris_setosa_,0.991359,0.00861,0.000031
5.4,3.4,1.5,0.4,b_Iris_setosa_,0.989595,0.010372,0.000033
5.2,3.4,1.4,0.2,b_Iris_setosa_,0.991675,0.008295,0.00003
4.3,3,1.1,0.1,b_Iris_setosa_,0.991961,0.008008,0.000031

```

5.3,3.7,1.5,0.2,b_Iris_setosa_,0.992273,0.007698,0.000029
 4.8,3,1.4,0.3,b_Iris_setosa_,0.988941,0.011024,0.000035
 4.6,3.4,1.4,0.3,b_Iris_setosa_,0.991559,0.00841,0.000031
 6,2.2,4,1,b_Iris_versicolor_,0.007382,0.989225,0.003393
 5.5,2.5,4,1.3,b_Iris_versicolor_,0.006417,0.989033,0.00455
 5.2,2.7,3.9,1.4,b_Iris_versicolor_,0.007379,0.988255,0.004366
 6.2,2.9,4.3,1.3,b_Iris_versicolor_,0.006957,0.989328,0.003715
 5.5,2.4,3.7,1,b_Iris_versicolor_,0.01306,0.984509,0.002431
 6.2,9,4.5,1.5,b_Iris_versicolor_,0.003952,0.985287,0.010761
 6.1,2.9,4.7,1.4,b_Iris_versicolor_,0.00407,0.985962,0.009968
 5.1,2.5,3,1.1,b_Iris_versicolor_,0.030698,0.967745,0.001556
 5.7,3,4.2,1.2,b_Iris_versicolor_,0.01178,0.985639,0.002581
 6.3,3.3,4.7,1.6,b_Iris_versicolor_,0.003894,0.984544,0.011562
 6.8,3,5.5,2.1,b_Iris_virginica_,0.000124,0.002832,0.997043
 6.9,3.1,5.4,2.1,b_Iris_virginica_,0.000143,0.004387,0.99547
 7.2,3.6,6.1,2.5,b_Iris_virginica_,0.00009,0.001484,0.998427
 6.5,3,5.2,2,b_Iris_virginica_,0.000197,0.009991,0.989812
 7.9,3.8,6.4,2,b_Iris_virginica_,0.000131,0.003799,0.99607
 6.4,2.8,5.6,2.1,b_Iris_virginica_,0.000109,0.001856,0.998035
 5.7,2.5,5,2,b_Iris_virginica_,0.000147,0.003295,0.996558
 6.2,3.4,5.4,2.3,b_Iris_virginica_,0.000123,0.002181,0.997695
 6.7,3,5.2,2.3,b_Iris_virginica_,0.000114,0.002315,0.99757
 4.9,2.5,4.5,1.7,b_Iris_virginica_,0.000769,0.284154,0.715077

Neuronų išėjimo reikšmių perskaičiavimas MS Excel programoje

Duomenų nukopijavimas į MS Excel programą ir normavimas

Normavimui naudojama formulė:

$$x_{ij} \leftarrow \frac{2x_{ij} - \min_{(x_{1j}, x_{2j}, \dots, x_{mj})} - \max_{(x_{1j}, x_{2j}, \dots, x_{mj})}}{\max_{(x_{1j}, x_{2j}, \dots, x_{mj})} - \min_{(x_{1j}, x_{2j}, \dots, x_{mj})}}$$

Normavimo formulės realizavimas MS Excel programoje:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	sepallength	sepalwidth	petallength	petalwidth	class	p_setosa	p_versicolor	p_virginica		Normalizuoti:	sepallength	sepalwidth	petallength	petalwidth	
2	4,8	3	1,4	0,1		0,99	0,01	0,00		\$A\$2:\$A\$31)	-0,849	-0,887	-1,000		
3	5,1	3,4	1,5	0,2		0,99	0,01	0,00			-0,556	0,500	-0,849	-0,917	
4	4,9	3,1	1,5	0,1		0,99	0,01	0,00			-0,667	0,125	-0,849	-1,000	
5	4,6	3,2	1,4	0,2		0,99	0,01	0,00			-0,833	0,250	-0,887	-0,917	
6	5,4	3,4	1,5	0,4		0,99	0,01	0,00			-0,389	0,500	-0,849	-0,750	
7	5,2	3,4	1,4	0,2		0,99	0,01	0,00			-0,500	0,500	-0,887	-0,917	
8	4,3	3	1,1	0,1		0,99	0,01	0,00			-1,000	0,000	-1,000	-1,000	
9	5,3	3,7	1,5	0,2		0,99	0,01	0,00			-0,444	0,875	-0,849	-0,917	
10	4,8	3	1,4	0,3		0,99	0,01	0,00			-0,722	0,000	-0,887	-0,833	
11	4,6	3,4	1,4	0,3		0,99	0,01	0,00			-0,833	0,500	-0,887	-0,833	
12	6,0	2,2	4	1		0,01	0,99	0,00			-0,056	-1,000	0,094	-0,250	
13	5,5	2,5	4	1,3		0,01	0,99	0,00			-0,333	-0,625	0,094	0,000	
14	5,2	2,7	3,9	1,4		0,01	0,99	0,00			-0,500	-0,375	0,057	0,083	
15	6,2	2,9	4,3	1,3		0,01	0,99	0,00			0,056	-0,125	0,208	0,000	
16	5,5	2,4	3,7	1		0,01	0,98	0,00			-0,333	-0,750	-0,019	-0,250	
17	6,0	2,9	4,5	1,5		0,00	0,99	0,01			-0,056	-0,125	0,283	0,167	
18	6,1	2,9	4,7	1,4		0,00	0,99	0,01			0,000	-0,125	0,358	0,083	
19	5,1	2,5	3	1,1		0,03	0,97	0,00			-0,556	-0,625	-0,283	-0,167	
20	5,7	3	4,2	1,2		0,01	0,99	0,00			-0,222	0,000	0,170	-0,083	
21	6,3	3,3	4,7	1,6		0,00	0,98	0,01			0,111	0,375	0,358	0,250	
22	6,8	3	5,5	2,1		0,00	0,00	1,00			0,389	0,000	0,660	0,667	
23	6,9	3,1	5,4	2,1		0,00	0,00	1,00			0,444	0,125	0,623	0,667	
24	7,2	3,6	6,1	2,5		0,00	0,00	1,00			0,611	0,750	0,887	1,000	
25	6,5	3	5,2	2		0,00	0,01	0,99			0,222	0,000	0,547	0,583	
26	7,9	3,8	6,4	2		0,00	0,00	1,00			1,000	1,000	1,000	0,583	
27	6,4	2,8	5,6	2,1		0,00	0,00	1,00			0,167	-0,250	0,698	0,667	
28	5,7	2,5	5	2		0,00	0,00	1,00			-0,222	-0,625	0,472	0,583	
29	6,2	3,4	5,4	2,3		0,00	0,00	1,00			0,056	0,500	0,623	0,833	
30	6,7	3	5,2	2,3		0,00	0,00	1,00			0,333	0,000	0,547	0,833	
31	4,9	2,5	4,5	1,7		0,00	0,28	0,72			-0,667	-0,625	0,283	0,333	
32															

11 pav. normavimo formulė MS Excel

Neuronų svorių lentelės

Svorai:	Node	0	1	2											
	Threshold	0,340	0,133	-5,340											
	3	4,877	-5,188	-5,196											
	4	-2,034	0,456	0,421											
	5	-1,641	-6,162	5,823											
	6	-1,659	-6,457	6,002											
	7	-4,704	5,522	-0,315											
	Node	3	4	5	6	7									
	Threshold	-1,686	-0,283	-5,234	-5,354	2,644									
	sepallength	-0,084	0,338	-0,909	-1,001	0,933									
	sepalwidth	2,097	-0,436	-2,167	-2,284	-2,081									
	petallength	-2,869	0,960	7,309	7,647	3,123									
	petalwidth	-2,933	1,095	6,167	6,295	3,129									

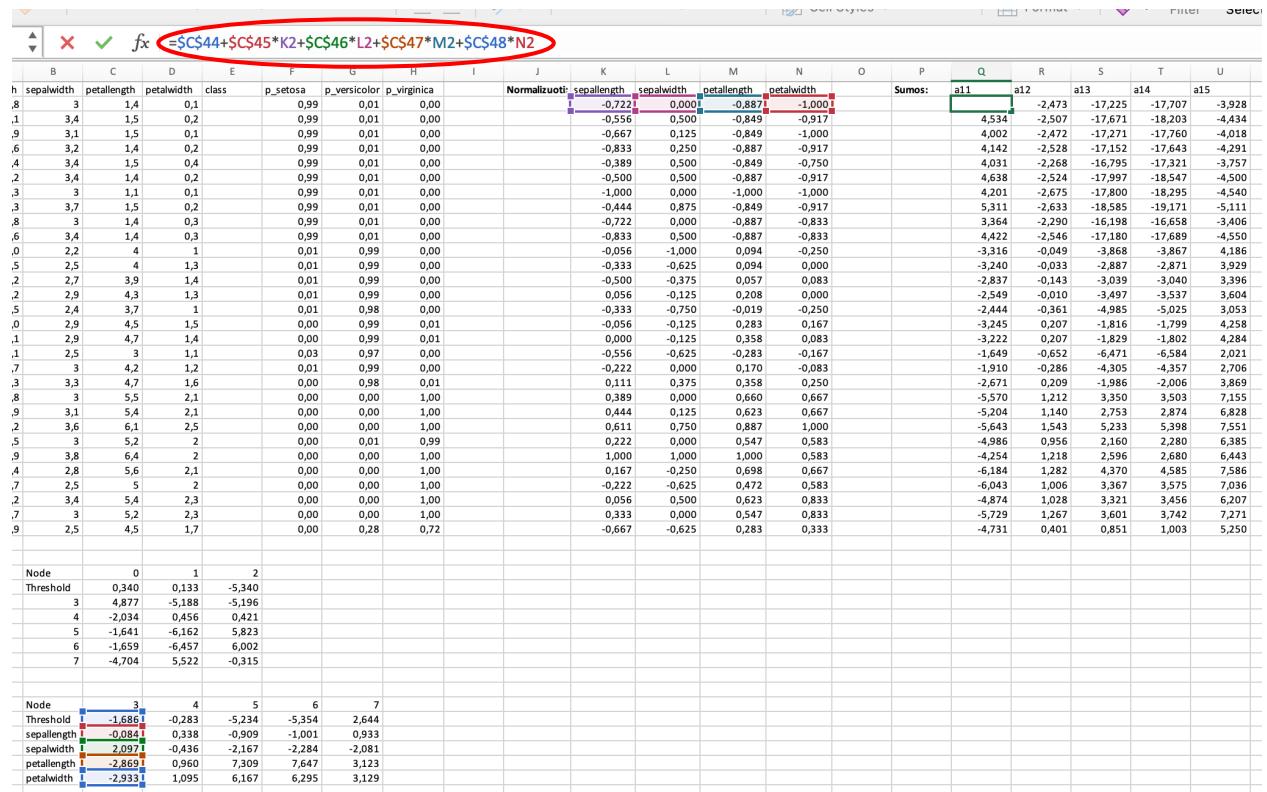
12 pav. neuronų svorių lentelės

Duomenų iėjimo vektorių ir paslėptų neuronų svorių vektorių sandaugų sumos

Naudojama sandaugos sumos formulė:

$$a_j = \sum_{k=0}^n w_{1k} x_k \quad j = 1, 2, 3, 4, 5.$$

Formulės realizavimas MS Excel programoje:



B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
h	sepalwidth	petallength	petalwidth	class	p_setosa	p_versicolor	p_virginica	Normalizuoti	sepallength	sepalwidth	petallength	petalwidth	Sumos:	a11	a12	a13	a14	a15	
,8	3	1,4	0,1		0,99	0,01	0,00		-0,722	0,000	-0,887	-1,000			-2,473	-17,225	-17,707	-3,928	
,1	3,4	1,5	0,2		0,99	0,01	0,00		-0,556	0,500	-0,849	-0,917		4,534	-2,507	-17,671	-18,203	-4,434	
,9	3,1	1,5	0,1		0,99	0,01	0,00		-0,667	0,125	-0,849	-1,000		4,002	-2,472	-17,271	-17,760	-4,018	
,6	3,2	1,4	0,2		0,99	0,01	0,00		-0,833	0,250	-0,887	-0,917		4,142	-2,528	-17,152	-17,643	-4,291	
,4	3,4	1,5	0,4		0,99	0,01	0,00		-0,389	0,500	-0,849	-0,750		4,031	-2,268	-16,795	-17,321	-3,757	
,2	3,4	1,4	0,2		0,99	0,01	0,00		-0,509	0,500	-0,887	-0,917		4,638	-2,524	-17,997	-18,547	-4,500	
,3	3	1,1	0,1		0,99	0,01	0,00		-1,000	0,000	-1,000	-1,000		4,201	-2,675	-17,800	-18,295	-4,540	
,3	3,7	1,5	0,2		0,99	0,01	0,00		-0,444	0,875	-0,849	-0,917		5,311	-2,633	-18,585	-19,171	-5,111	
,8	3	1,4	0,3		0,99	0,01	0,00		-0,722	0,000	-0,887	-0,833		3,364	-2,290	-16,198	-16,658	-3,406	
,6	3,4	1,4	0,3		0,99	0,01	0,00		-0,833	0,500	-0,887	-0,833		4,422	-2,546	-17,180	-17,689	-4,550	
,0	2,2	4	1		0,01	0,99	0,00		-0,056	-1,000	0,094	-0,250		-3,316	-0,049	-3,868	-3,867	4,186	
,5	2,5	4	1,3		0,01	0,99	0,00		-0,333	-0,625	0,094	0,000		-3,240	-0,033	-2,887	-2,871	3,929	
,2	2,7	3,9	1,4		0,01	0,99	0,00		-0,500	-0,375	0,057	0,083		-2,837	-0,143	-3,039	-3,040	3,396	
,2	2,9	4,3	1,3		0,01	0,99	0,00		0,056	-0,125	0,208	0,000		-2,549	-0,010	-3,497	-3,537	3,604	
,5	2,4	3,7	1		0,01	0,98	0,00		-0,333	-0,750	-0,019	-0,250		-2,444	-0,361	-4,985	-5,025	3,053	
,0	2,9	4,5	1,5		0,00	0,99	0,01		-0,056	-0,125	0,283	0,167		-3,245	0,207	-1,816	-1,799	4,258	
,1	2,9	4,7	1,4		0,00	0,99	0,01		0,000	-0,125	0,358	0,083		-3,222	0,207	-1,829	-1,802	4,284	
,1	2,5	3	1,1		0,03	0,97	0,00		-0,556	-0,625	-0,283	-0,167		-1,649	-0,652	-6,471	-6,584	2,021	
,7	3	4,2	1,2		0,01	0,99	0,00		-0,222	0,000	0,170	-0,083		-1,910	-0,286	-4,305	-4,357	2,706	
,3	3,3	4,7	1,6		0,00	0,98	0,01		0,111	0,375	0,358	0,250		-2,671	0,209	-1,986	-2,006	3,869	
,8	3	5,5	2,1		0,00	0,00	1,00		0,389	0,000	0,660	0,667		-5,570	1,212	3,350	3,503	7,155	
,9	3,1	5,4	2,1		0,00	0,00	1,00		0,444	0,125	0,623	0,667		-5,204	1,140	2,753	2,874	6,828	
,2	3,6	6,1	2,5		0,00	0,00	1,00		0,611	0,750	0,887	1,000		-5,643	1,543	5,233	5,398	7,551	
,5	3	5,2	2		0,00	0,01	0,99		0,222	0,000	0,547	0,583		-4,986	0,956	2,160	2,280	6,385	
,9	3,8	6,4	2		0,00	0,00	1,00		1,000	1,000	1,000	0,583		-4,254	1,218	2,596	2,680	6,443	
,4	2,8	5,6	2,1		0,00	0,00	1,00		0,167	-0,250	0,698	0,667		-6,184	1,282	4,370	4,585	7,586	
,7	2,5	5	2		0,00	0,00	1,00		-0,222	-0,625	0,472	0,583		-6,043	1,006	3,367	3,575	7,036	
,2	3,4	5,4	2,3		0,00	0,00	1,00		0,056	0,500	0,623	0,833		-4,874	1,028	3,321	3,456	6,207	
,7	3	5,2	2,3		0,00	0,00	1,00		0,333	0,000	0,547	0,833		-5,729	1,267	3,601	3,742	7,271	
,9	2,5	4,5	1,7		0,00	0,28	0,72		-0,667	-0,625	0,283	0,333		-4,731	0,401	0,851	1,003	5,250	
Node	0	1	2																
Threshold	0,340	0,133	-5,340																
3	4,877	-5,188	-5,196																
4	-2,034	0,456	0,421																
5	-1,641	-6,162	5,823																
6	-1,659	-6,457	6,002																
7	-4,704	5,522	-0,315																
Node	3	4	5	6	7														
Threshold	-1,686	-0,283	-5,234	-5,354	2,644														
sepallength	-0,084	0,338	-0,909	-1,001	0,933														
sepalwidth	2,097	-0,436	-2,167	-2,284	-2,081														
petallength	-2,869	0,960	7,309	7,647	3,123														
petalwidth	-2,933	1,095	6,167	6,295	3,129														

13 pav. sandaugos sumos formulės realizavimas MS Excel

l	O	P	Q	R	S	T	U	V	W	X
dth		Sumos:	a11	a12	a13	a14	a15			
-1,000			3,853	-2,473	-17,225	-17,707	-3,928			
-0,917			4,534	-2,507	-17,671	-18,203	-4,434			
-1,000			4,002	-2,472	-17,271	-17,760	-4,018			
-0,917			4,142	-2,528	-17,152	-17,643	-4,291			
-0,750			4,031	-2,268	-16,795	-17,321	-3,757			
-0,917			4,638	-2,524	-17,997	-18,547	-4,500			
-1,000			4,201	-2,675	-17,800	-18,295	-4,540			
-0,917			5,311	-2,633	-18,585	-19,171	-5,111			
-0,833			3,364	-2,290	-16,198	-16,658	-3,406			
-0,833			4,422	-2,546	-17,180	-17,689	-4,550			
-0,250			-3,316	-0,049	-3,868	-3,867	4,186			
0,000			-3,240	-0,033	-2,887	-2,871	3,929			
0,083			-2,837	-0,143	-3,039	-3,040	3,396			
0,000			-2,549	-0,010	-3,497	-3,537	3,604			
-0,250			-2,444	-0,361	-4,985	-5,025	3,053			
0,167			-3,245	0,207	-1,816	-1,799	4,258			
0,083			-3,222	0,207	-1,829	-1,802	4,284			
-0,167			-1,649	-0,652	-6,471	-6,584	2,021			
-0,083			-1,910	-0,286	-4,305	-4,357	2,706			
0,250			-2,671	0,209	-1,986	-2,006	3,869			
0,667			-5,570	1,212	3,350	3,503	7,155			
0,667			-5,204	1,140	2,753	2,874	6,828			
1,000			-5,643	1,543	5,233	5,398	7,551			
0,583			-4,986	0,956	2,160	2,280	6,385			
0,583			-4,254	1,218	2,596	2,680	6,443			
0,667			-6,184	1,282	4,370	4,585	7,586			
0,583			-6,043	1,006	3,367	3,575	7,036			
0,833			-4,874	1,028	3,321	3,456	6,207			
0,833			-5,729	1,267	3,601	3,742	7,271			
0,333			-4,731	0,401	0,851	1,003	5,250			

14 pav. gauti rezultatai, naudojant sandaugos sumos formulę

Sigmoidinės funkcijos reikšmių skaičiavimas nuo gautų sumų

Naudojama formulė:

$$f(a_j) = \frac{1}{1 + e^{-a_j}} \quad j = 1, 2, 3, 4, 5.$$

Formulės realizavimas MS Excel programoje:

L	M	N	O	P	Q	R	S	T	U	V	W
sepalwidth	petallength	petalwidth		Sumos:	a11	a12	a13	a14	a15		
0,000	-0,887	-1,000			3,853	-2,473	-17,225	-17,707	-3,928		
0,500	-0,849	-0,917			4,534	-2,507	-17,671	-18,203	-4,434		
0,125	-0,849	-1,000			4,002	-2,472	-17,271	-17,760	-4,018		
0,250	-0,887	-0,917			4,142	-2,528	-17,152	-17,643	-4,291		
0,500	-0,849	-0,750			4,031	-2,268	-16,795	-17,321	-3,757		
0,500	-0,887	-0,917			4,638	-2,524	-17,997	-18,547	-4,500		
0,000	-1,000	-1,000			4,201	-2,675	-17,800	-18,295	-4,540		
0,875	-0,849	-0,917			5,311	-2,633	-18,585	-19,171	-5,111		
0,000	-0,887	-0,833			3,364	-2,290	-16,198	-16,658	-3,406		
0,500	-0,887	-0,833			4,422	-2,546	-17,180	-17,689	-4,550		
-1,000	0,094	-0,250			3,316	-0,049	-3,868	-3,867	4,186		
-0,625	0,094	0,000			3,240	-0,033	-2,887	-2,871	3,929		
-0,375	0,057	0,083			2,837	-0,143	-3,039	-3,040	3,396		
-0,125	0,208	0,000			2,549	-0,010	-3,497	-3,537	3,604		
-0,750	-0,019	-0,250			2,444	-0,361	-4,985	-5,025	3,053		
-0,125	0,283	0,167			3,245	0,207	-1,816	-1,799	4,258		
-0,125	0,358	0,083			3,222	0,207	-1,829	-1,802	4,284		
-0,625	-0,283	-0,167			1,649	-0,652	-6,471	-6,584	2,021		
0,000	0,170	-0,083			1,910	-0,286	-4,305	-4,357	2,706		
0,375	0,358	0,250			2,671	0,209	-1,986	-2,006	3,869		
0,000	0,660	0,667			5,570	1,212	3,350	3,503	7,155		
0,125	0,623	0,667			5,204	1,140	2,753	2,874	6,828		
0,750	0,887	1,000			5,643	1,543	5,233	5,398	7,551		
0,000	0,547	0,583			4,986	0,956	2,160	2,280	6,385		
1,000	1,000	0,583			4,254	1,218	2,596	2,680	6,443		
-0,250	0,698	0,667			6,184	1,282	4,370	4,585	7,586		
-0,625	0,472	0,583			6,043	1,006	3,367	3,575	7,036		
0,500	0,623	0,833			4,874	1,028	3,321	3,456	6,207		
0,000	0,547	0,833			5,729	1,267	3,601	3,742	7,271		
-0,625	0,283	0,333			4,731	0,401	0,851	1,003	5,250		
				Sumų aktyvacija fa11	fa12	fa13	fa14	fa15			
				Q2))	0,078	0,000	0,000	0,019			
					0,989	0,075	0,000	0,000	0,012		
					0,982	0,078	0,000	0,000	0,018		
					0,984	0,074	0,000	0,000	0,014		
					0,983	0,094	0,000	0,000	0,023		
				Sumų aktyvacija fa11	fa12	fa13	fa14	fa15			
					0,979	0,078	0,000	0,000	0,019		
					0,989	0,075	0,000	0,000	0,012		
					0,982	0,078	0,000	0,000	0,018		
					0,984	0,074	0,000	0,000	0,014		
					0,985	0,064	0,000	0,000	0,011		
					0,995	0,067	0,000	0,000	0,006		
					0,967	0,092	0,000	0,000	0,032		
					0,988	0,073	0,000	0,000	0,010		
					0,035	0,488	0,020	0,021	0,985		
					0,038	0,492	0,053	0,054	0,981		
					0,055	0,464	0,046	0,046	0,968		
					0,073	0,497	0,029	0,028	0,974		
					0,080	0,411	0,007	0,007	0,955		
					0,038	0,552	0,140	0,142	0,986		
					0,038	0,552	0,138	0,142	0,986		
					0,161	0,342	0,002	0,001	0,883		
					0,129	0,429	0,013	0,013	0,937		
					0,065	0,552	0,121	0,119	0,980		
					0,004	0,771	0,966	0,971	0,999		
					0,005	0,758	0,940	0,947	0,999		
					0,004	0,824	0,995	0,995	0,999		
					0,007	0,722	0,897	0,907	0,998		
					0,014	0,772	0,931	0,936	0,998		
					0,002	0,783	0,988	0,990	0,999		
					0,002	0,732	0,967	0,973	0,999		
					0,008	0,737	0,965	0,969	0,998		
					0,003	0,780	0,973	0,977	0,999		
					0,009	0,599	0,701	0,732	0,995		

15 pav. sigmoidinės f-jos MS Excel ir gauti rezultatai

Sigmoidinės funkcijos gautų reikšmių (t.y. paslėpto sluoksnio išėjimų) vektorių ir paslėptų neuronų svorių vektorių sandaugų sumų skaičiavimas

16 pav. sandaugų sumų skaičiavimas MS Excel ir gauti rezultatai

Tikimybės

17 pav. tikimybių skaičiavimas MS Excel ir gauti rezultatai

Gautų tikimybių palyginimas su neuroninio tinklo gautomis tikimybėmis

18 pav. gautų tikimybių palyginimas su neuroninio tinklo gautomis tikimybėmis

IŠVADOS

Perskaičiavus svorius su MS Excel programa, gavome tokius pat rezultatus, lyginant su programos WEKA rezultatais.

Keičiant mokymosi greitį ir momentum reikšmes, pastebėta, kad mažesnės learning rate ir momentum reikšmės duoda geresnius rezultatus, taip pat pastebėta, kad didesni neuronų sluoksnių kiekiai nebūtinai reiškia geresnius rezultatus.