

Ataskaita

Laimonas Beniušis, 1410102

4 Užduotis

Duomenų klasifikavimas neuroniniais tinklais WEKA sistemoje

Analizei naudojami duomenys - Irisų (iris.arff) byla.

Duomenų klasės:

1. Iris-setosa
2. Iris-versicolor
3. Iris-virginica

Iš bylos iris.arff duomenys išskaidomi į 2 atskiras bylas (irisTrain.arff, irisEvaluate.arff).

Iš kiekvienos duomenų klasės perkeliama 10 vienos klasės duomenų įrašų į atskirą bylą.

Dabar irisTrain.arff byla turi 120 įrašų, o irisEvaluate.arff turi 30 įrašų.

1. Etapas

Sudaromos mokslinio darbo sekos.

Visi naudojami darbo sekos komponentai:

ArffLoader – pakrauna .arff tipo bylą su duomenimis

ArffSaver – išsaugo rezultatus .arff byloje

ClassAssigner – paruošia duomenis klasifikavimui

ClassifierPerformanceEvaluator – patikrina klasifikatoriaus rezultatus

CrossValidationFoldMaker – suskirsto duomenis į N lygių aibių. Kiekvienoje aibėje yra 2 duomenų grupės: mokymo, testavimo (sątykis 9/1). Tada tarp N aibių patikrina rezultatus tarpusavyje galutinį rezultatą gauna iš jų vidurkio.

MultilayerPerceptron – neuronų tinklas, kuriuo klasifikuojami duomenys. Šiuo atveju, tai yra daugiasluoksnis perceptronas.

PredictionAppender – klasifikatorius, kuris gauna testavimo duomenų rezultatus, kurie nebuvo naudojami mokymui

SerializedModelSaver - neuronų tinklui išsaugoti

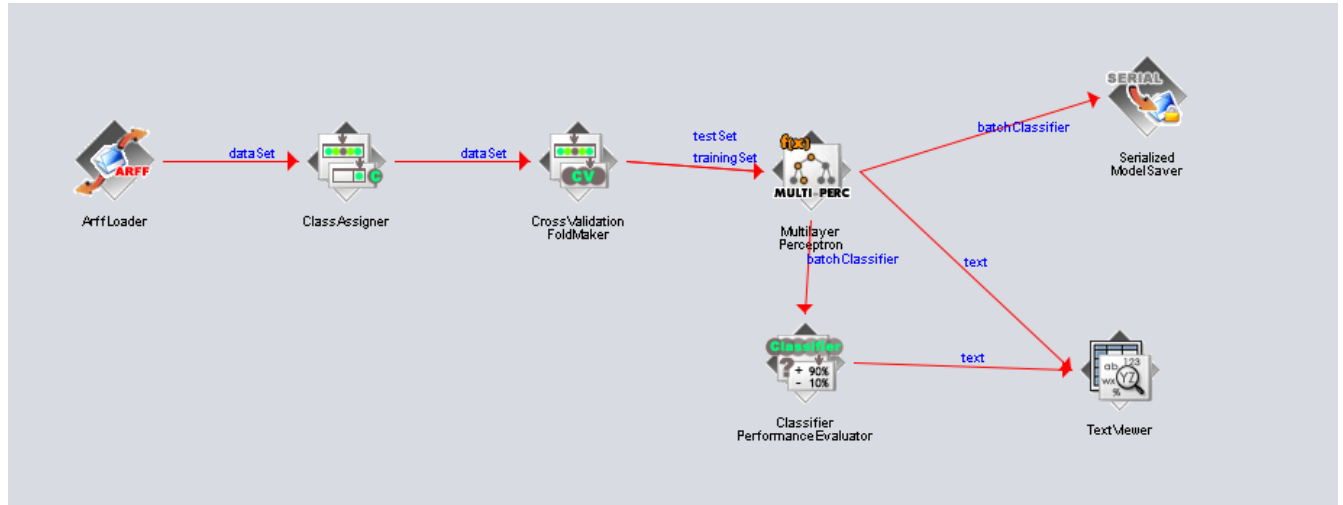
TextViewer – rezultatams peržiūrėti

TextSaver – išsaugo rezultatus į teksto bylą

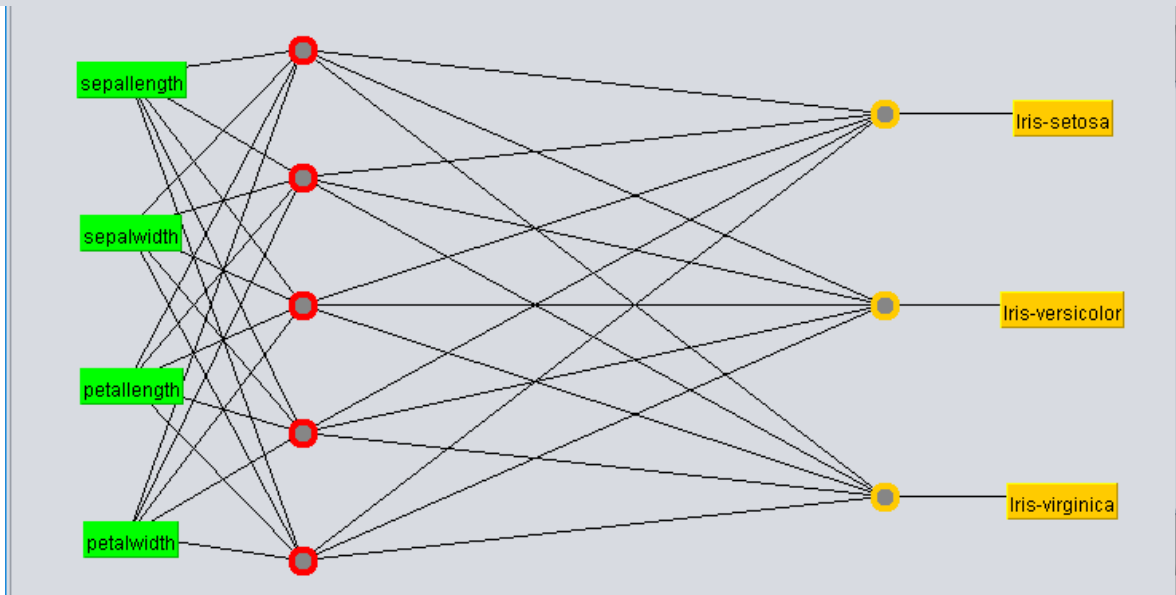
TestSetMaker – iš duomenų padaro testavimo aibę

TrainingSetMaker – iš duomenų padaro mokymo aibę

Seka 1



Visose sekose
naudotas neuronų
tinklas:



Mokymo greitis (learning rate) 0.1

Atsparumas krypties pokyčiui (momentum) 0.1

Epochų skaičius: 500

Seed = 1337

Naudojami irisTrain.arff duomenys.

Geri rezultatai (96.6% tikslumo) buvo pasiekti su pakankamai mažu (5) paslėptų neuronų kiekiu.

Idealaus rezultato (100%) tikslumo pasiekti nepavyko.

Žemiau pateikiami pilni TextViewer duomenys.

=== Evaluation result ===

Scheme: MultilayerPerceptron

Options: -L 0.1 -M 0.1 -N 500 -V 0 -S 1337 -E 20 -H 5 -R

Relation: iris

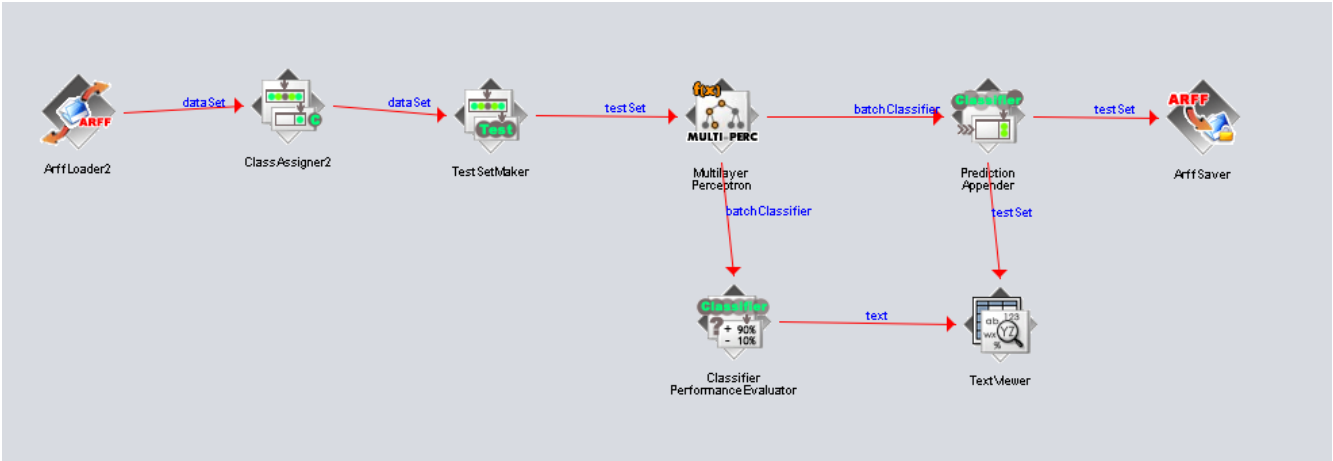
Correctly Classified Instances	116	96.6667 %
Incorrectly Classified Instances	4	3.3333 %
Kappa statistic	0.95	
Mean absolute error	0.0482	
Root mean squared error	0.1405	
Relative absolute error	10.8394 %	
Root relative squared error	29.8002 %	
Total Number of Instances	120	

=== 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		Iris-setosa
	0.925	0.013	0.974	0.925	0.949	0.925	0.993	0.987	Iris-versicolor
	0.975	0.038	0.929	0.975	0.951	0.927	0.993	0.986	Iris-virginica
Weighted Avg.	0.967	0.017	0.967	0.967	0.967	0.950	0.995	0.991	

=== Confusion Matrix ===

```
a b c <-- classified as
40 0 0 | a = Iris-setosa
0 37 3 | b = Iris-versicolor
0 1 39 | c = Iris-virginica
```



Naudojamas anksčiau apmokytas neuronų tinklo modelis.
Testavimui naudojami likę duomenys (irisEvaluate.arff).
Gautas tikslumas 100%.

=== Evaluation result ===

Scheme: MultilayerPerceptron
Options: -L 0.1 -M 0.1 -N 500 -V 0 -S 1337 -E 20 -H 5 -R
Relation: iris

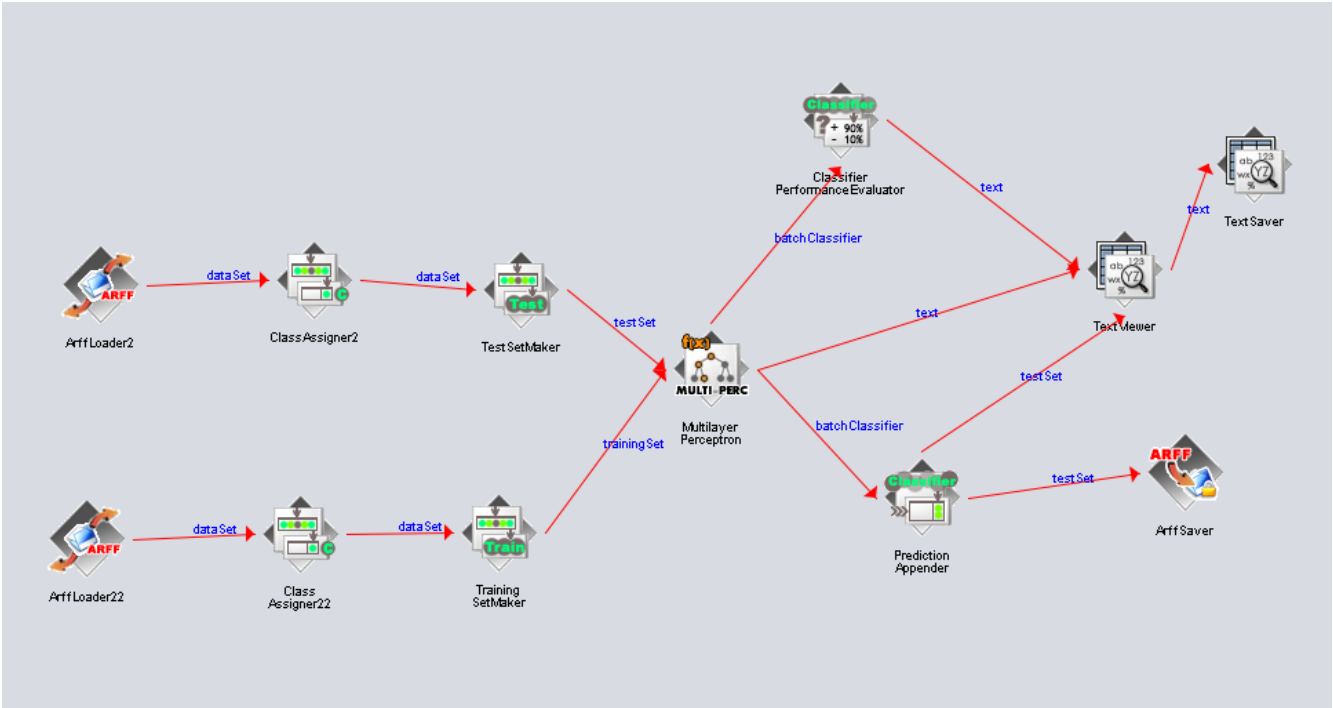
Correctly Classified Instances	30	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0247		
Root mean squared error	0.0647		
Total Number of Instances	30		

=== 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	Iris-setosa
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-versicolor
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-virginica
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a b c <-- classified as
10 0 0 | a = Iris-setosa
0 10 0 | b = Iris-versicolor
0 0 10 | c = Iris-virginica



Atskirai pakraunamos mokymo ir testavimo (nežinomų) duomenų aibės. Rezultatai išsaugomi. Gautas 100% tikslumas.

=== Evaluation result ===

Scheme: MultilayerPerceptron
Options: -L 0.1 -M 0.1 -N 500 -V 0 -S 1337 -E 20 -H 5 -R
Relation: iris

Correctly Classified Instances	30	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.022		
Root mean squared error	0.0461		
Relative absolute error	4.9411 %		
Root relative squared error	9.7822 %		
Total Number of Instances	30		

=== 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	Iris-setosa
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-versicolor
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-virginica
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a b c <-- classified as
10 0 0 | a = Iris-setosa
0 10 0 | b = Iris-versicolor
0 0 10 | c = Iris-virginica

2. Etapas

Sudaromas neuronų tinklas skaičiuoklėje (spreadsheet).

Rezultatų palyginimas

Wekos reikšmės			Skaičiuoklės reikšmės			Skirtumas		
0.974235	0.025746	1.8E-05	0.982775061	0.0248724773226	1.720763E-05	0.0085400608	0.0008735227	7.923695E-07
0.970488	0.029492	2E-05	0.980202446	0.0279847310367	1.8987234E-05	0.0097144459	0.001507269	1.012766E-06
0.973189	0.026793	1.9E-05	0.981991984	0.0259424953025	1.7775882E-05	0.0088029843	0.0008505047	1.224118E-06
0.972081	0.027899	2E-05	0.980635906	0.0266150966623	1.8718718E-05	0.0085549061	0.0012839033	1.281282E-06
0.97478	0.025202	1.8E-05	0.982979492	0.024502331453	1.7062004E-05	0.0081994921	0.0006996685	9.379956E-07
0.975608	0.024373	1.9E-05	0.981691545	0.0226478248159	1.7927891E-05	0.006083545	0.0017251752	1.072109E-06
0.974185	0.025796	1.9E-05	0.981810185	0.0247755553338	1.7899883E-05	0.007625185	0.0010204447	1.100117E-06
0.973678	0.026304	1.9E-05	0.982213992	0.0251837493042	1.7606249E-05	0.0085359919	0.0011202507	1.393751E-06
0.970872	0.029107	2.1E-05	0.979375863	0.0280171618129	1.9582018E-05	0.0085038631	0.0010898382	1.417982E-06
0.972024	0.027957	1.9E-05	0.981515709	0.0268890234197	1.809911E-05	0.0094917095	0.0010679766	9.008901E-07
0.013732	0.97797	0.008298	0.016353642	0.9750149574942	0.0067182828	0.0026216416	0.0029550425	0.0015797172
0.010798	0.977458	0.011744	0.013042424	0.9689546066107	0.00912269759	0.0022444239	0.0085033934	0.0026213024
0.006137	0.96369	0.030173	0.006497745	0.9424858121646	0.02667115392	0.0003607448	0.0212041878	0.0035018461
0.007614	0.976854	0.015532	0.004847492	0.9629228610975	0.02811593501	0.0027665084	0.0139311389	0.012583935
0.005126	0.958923	0.035951	0.004425187	0.9294938129598	0.04468833904	0.0007008131	0.029429187	0.008737339
0.00863	0.976193	0.015177	0.007325843	0.9619853005734	0.0186330857	0.0013041567	0.0142076994	0.0034560857
0.006408	0.96388	0.029713	0.008073587	0.945788526819	0.01993947787	0.0016655873	0.0180914732	0.0097735221
0.051989	0.945936	0.002075	0.037201386	0.9814583297126	0.00289444661	0.0147876144	0.0355223297	0.0008194466
0.012019	0.97852	0.00946	0.011492195	0.9736002379582	0.00992163206	0.0005268054	0.004919762	0.0004616321
0.010795	0.978413	0.010792	0.009143704	0.9700751183847	0.01280570048	0.0016512956	0.0083378816	0.0020137005
0.000105	0.007451	0.992444	0.000104982	0.0072825763666	0.99656502749	1.787447E-08	0.0001684236	0.0041210275
0.000318	0.03237	0.967312	0.000259872	0.0225894079516	0.98678725446	5.812757E-05	0.009780592	0.0194752545
0.00018	0.011534	0.988286	0.000167575	0.0105683540384	0.99413941771	1.242468E-05	0.000965646	0.0058534177
0.000336	0.029944	0.96972	0.00028661	0.0221408456516	0.98548451844	4.938982E-05	0.0078031543	0.0157645184
0.000147	0.009769	0.990084	0.000137871	0.009177601231	0.9954028767	9.128974E-06	0.0005913988	0.0053188767
0.000127	0.008601	0.991272	0.000117104	0.0081224811991	0.99616762225	9.896375E-06	0.0004785188	0.0048956223
0.000758	0.276585	0.722657	0.00053162	0.1362046628238	0.90685696443	0.0002263803	0.1403803372	0.1841999644
0.00022	0.014325	0.985455	0.000189301	0.0120080388809	0.99321919679	3.06995E-05	0.0023169611	0.0077641968
0.000214	0.016096	0.983689	0.000169477	0.0125113142153	0.99408114604	4.452282E-05	0.0035846858	0.010392146
0.000129	0.007978	0.991893	0.000145064	0.0080775186643	0.99518533866	1.606399E-05	9.951866E-05	0.0032923387

Gauti svoriai, ir visi tarpiniai skaičiavimai yra atskiroje skaičiuoklės byloje.

Dekarto koordinatų plokštumos vaizdas

