

Data Mining: STATISTICA



The University of Iowa

Intelligent Systems Laboratory

Outline

- Prepare the data
- Classification and regression

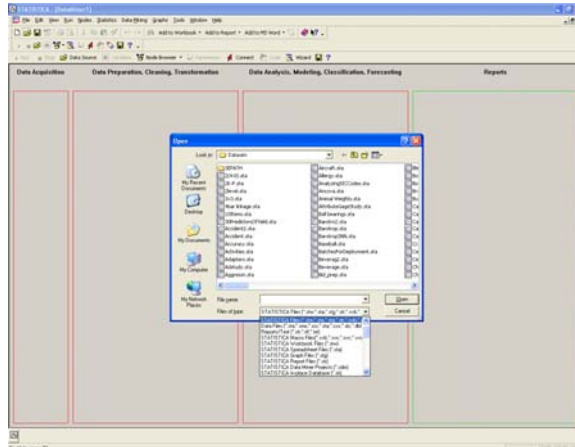


The University of Iowa

Intelligent Systems Laboratory

Prepare the Data

- Statistica can read from Excel, .txt and many other types of files
- Compared with WEKA, Statistica is much easier in terms of data preparing

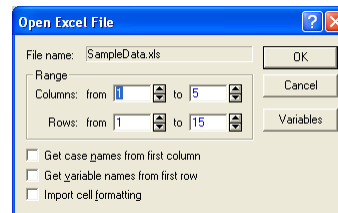
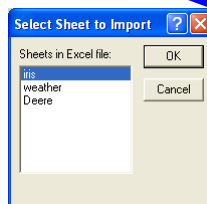
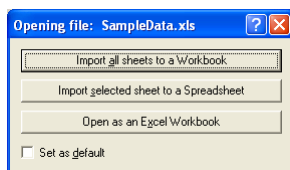


The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Click the “Import selected sheet to Spreadsheet”
- Select the desired Excel sheet where your data is stored
- Get variable names from the first row



The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Change variable type

	1 outlook	2 temperature	3 humidity	4 windy	5 play
1	sunny	hot	high	0	no
2	sunny	hot	high	1	no
3	overcast	hot	high	0	yes
4	rainy	mild	high	0	yes
5	rainy	cool	normal	0	yes
6	rainy	cool	normal	1	no
7	overcast	cool	normal	1	yes
8	sunny	mild	high	0	no
9	sunny	cool	normal	0	yes
10	rainy	mild	normal	0	yes
11	sunny	mild	normal	1	yes
12	overcast	mild	high	1	yes
13	overcast	hot	normal	0	yes
14	rainy	mild	high	1	no



The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Change variable type

Variable 1

Name: outlook Type: Text OK

Measurement Type: Auto Length: 8 Cancel

Excluded Label Case State MD code: 99999999 << >>

Display format

- General
- Number
- Date
- Time
- Scientific
- Currency
- Percentage
- Fraction
- Custom

All Specs...
Text Labels...
Values/Stats...
Properties...
[Bundles]...

Long name (label or formula with Functions): Function guide

Labels: use any text. Formulas: use variable names or v1, v2, ..., v0 is case #.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after)



The University of Iowa

Intelligent Systems Laboratory

Classification and Regression

- C&RT
- Boosting tree
- Neural Networks



The University of Iowa

Intelligent Systems Laboratory

C&RT Classification

- Iris data is used as an example data set

	1	2	3	4	5
	sepalength	sepalwidth	petallength	petalwidth	class
1	5.1	3.5	1.4	0.2	iris-setosa
2	4.9	3	1.4	0.2	iris-setosa
3	4.7	3.2	1.3	0.2	iris-setosa
4	4.6	3.1	1.5	0.2	iris-setosa
5	5	3.6	1.4	0.2	iris-setosa
6	5.4	3.9	1.7	0.4	iris-setosa
7	4.6	3.4	1.4	0.3	iris-setosa
8	5	3.4	1.5	0.2	iris-setosa
9	4.4	2.9	1.4	0.2	iris-setosa
10	4.9	3.1	1.5	0.1	iris-setosa
11	5.4	3.7	1.5	0.2	iris-setosa
12	4.8	3.4	1.6	0.2	iris-setosa
13	4.8	3	1.4	0.1	iris-setosa
14	4.3	3	1.1	0.1	iris-setosa
15	5.8	4	1.2	0.2	iris-setosa
16	5.7	4.4	1.5	0.4	iris-setosa
17	5.4	3.9	1.3	0.4	iris-setosa
18	5.1	3.5	1.4	0.3	iris-setosa
19	5.7	3.8	1.7	0.3	iris-setosa
20	5.1	3.8	1.5	0.3	iris-setosa
21	5.4	3.4	1.7	0.2	iris-setosa
22	5.1	3.7	1.5	0.4	iris-setosa
23	4.6	3.6	1	0.2	iris-setosa
24	5.1	3.3	1.7	0.5	iris-setosa
25	4.8	3.4	1.9	0.2	iris-setosa
26	5	3	1.6	0.2	iris-setosa



The University of Iowa

Intelligent Systems Laboratory

C&RT Classification

- Click "Data Mining" menu and find the "Interactive Trees"

The screenshots illustrate the initial setup for C&RT classification in SPSS:

- Top Left:** The SPSS menu bar with "Data Mining" highlighted, and the "Interactive Trees" option selected under the "C&RT" sub-menu.
- Top Right:** The "General Classification and Regression Trees: SampleData" dialog box. "Standard C&RT" is selected under "Type of analysis", and "Quick: specs dialog" is chosen under "Specification method".
- Middle Left:** The "Standard C&RT: SampleData" dialog box. "Categorical response (categorical dependent variable)" is selected under "Variables".
- Middle Right:** The "Select dependent (cat, categorical, and continuous predictors)" dialog box, showing the selection of "class" as the dependent variable.
- Bottom Left:** The "Standard C&RT: SampleData" dialog box with "Response codes" set to "none".
- Bottom Right:** The "Select codes for the categorical response" dialog box, showing the selection of "class" and the codes "Iris-setosa" and "Iris-virginica".



The University of Iowa

Intelligent Systems Laboratory

C&RT Classification

- View the final tree and understand the results

The screenshots show the final results of the C&RT classification:

- Left Panel:** The "GC&RT Results: SampleData" window. The "Tree view" tab is active, showing various options for viewing the tree (e.g., Tree browser, Tree graph, Tree layout, Tree structure, Terminal nodes, Tree sequence, Cost sequence, Importance, Importance, Design terms, V-fold crossvalidation & tree sequence).
- Right Panel:** A "Tree 1 graph for class" showing the decision tree structure. The tree is rooted at the top with a split on "petalwidth". The left branch (petalwidth <= 0.800000) leads to a node for "Iris-setosa" (N=150). The right branch (petalwidth > 0.800000) leads to a node for "Iris-versicolor" (N=100). Further splits on "petalwidth" and "petalength" lead to terminal nodes for "Iris-versicolor" and "Iris-virginica".

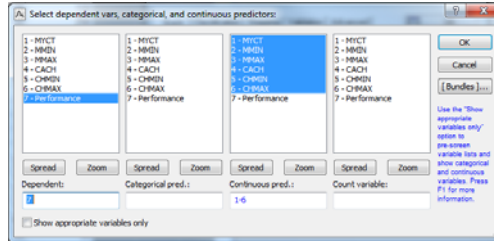
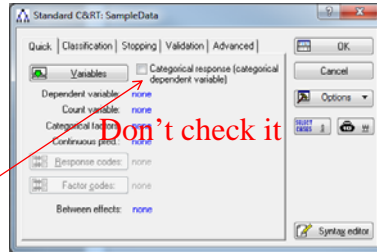
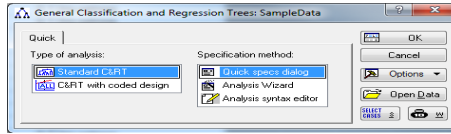
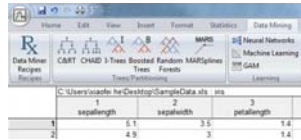


The University of Iowa

Intelligent Systems Laboratory

C&RT---Regression

- Use the CPU data set and select the regression analysis

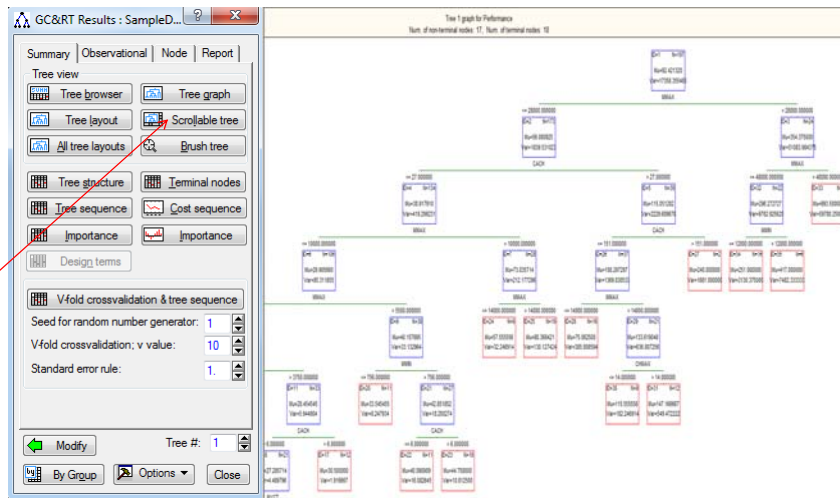


The University of Iowa

Intelligent Systems Laboratory

C&RT---Regression

- Regression tree structure



The University of Iowa

Intelligent Systems Laboratory

C&RT---Regression

Predicted values

GC&RT Results : SampleD...
 Summary Observational Node | Report
 Sample
 Analysis Test set
 Prediction Surrogate
 Predicted values
 Save predicted values
 Predicted vs. residuals
 Observed vs. predicted
 Observed vs residuals
 Probability plot of residuals
 Histogram of Residuals
 Modify Tree #: 1
 By Group Options Close

Workbook*
 General Classification/Re...
 Classification and Reg...
 Tree 1 graph for P...
 Predicted values 1

Predicted values 1 (SampleData)
 Dependent variable: Performance
 Options: Continuous response, Tree number 1, Analysis sample

Observed value	Predicted value	Standard error	Terminal node
193 000	245 000	41 000	27 000
253 000	251 000	55 949	34 000
253 000	251 000	55 949	34 000
253 000	251 000	55 949	34 000
132 000	147 167	23 440	31 000
290 000	251 000	55 949	34 000
381 000	417 000	86 505	35 000
381 000	417 000	86 505	35 000
749 000	993 500	244 500	33 000
1238 000	993 500	244 500	33 000
23 000	23 333	0 942	15 000
24 000	23 333	0 942	15 000
70 000	75 062	19 620	28 000
117 000	115 556	13 499	30 000
15 000	17 667	1 312	12 000
64 000	80 364	11 407	25 000
23 000	20 767	1 130	14 000
29 000	29 000	1 870	18 000
22 000	20 767	1 130	14 000
124 000	75 062	19 620	28 000
35 000	40 090	4 010	22 000
39 000	40 090	4 010	22 000
40 000	40 090	4 010	22 000
45 000	40 090	4 010	22 000
28 000	33 545	2 871	20 000
21 000	20 767	1 130	14 000
27 000	33 545	2 871	20 000
22 000	23 333	0 942	15 000
28 000	26 230	1 476	19 000
27 000	26 230	1 476	19 000
102 000	75 062	19 620	28 000
102 000	75 062	19 620	28 000
74 000	74 062	19 620	28 000

The University of Iowa Intelligent Systems Laboratory

Boosting tree Classification

- In "Data Mining" menu and find the "Boosted Trees"

Home Edit View Insert Format Statistics Data Mining
 Data Miner Recipes
 C&RT CHAD I-Trees Boosted Random MARS
 Trees Forests
 Machine Learning
 GAM Learning

Boosted Trees: SampleData
 Quick |
 Type of analysis:
 Classification Analysis
 Regression Analysis
 OK Cancel Options Open Data

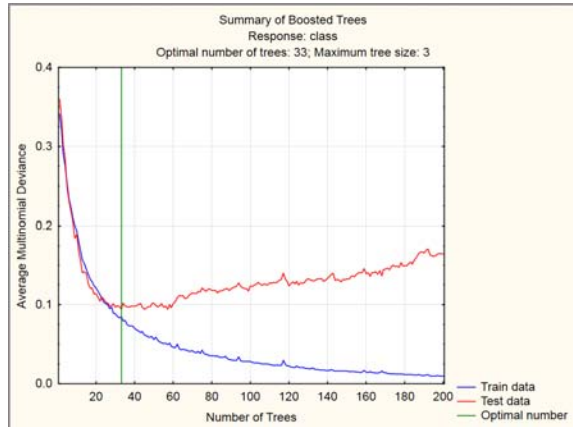
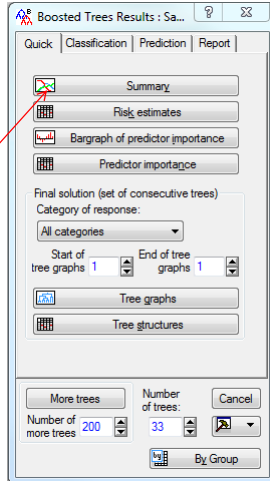
Boosted Trees Specifications: SampleData
 Quick | Classification | Advanced |
 Variables
 Dependent variable: none
 Count variable: none
 Categorical factors: none
 Correlates: none
 Response codes: none
 Factor codes: none

Select dependent vars, categorical, and continuous predictors
 1 - sepalwidth
 2 - sepalwidth
 3 - petalwidth
 4 - petalwidth
 5 - class
 1 - sepalwidth
 2 - sepalwidth
 3 - petalwidth
 4 - petalwidth
 5 - class
 1 - sepalwidth
 2 - sepalwidth
 3 - petalwidth
 4 - petalwidth
 5 - class
 Dependents: Categorical pred.: 1-4 Continuous pred.: Count variable: 5
 Show appropriate variables only

The University of Iowa Intelligent Systems Laboratory

Boosting tree Classification

- See the results and predictor's importance

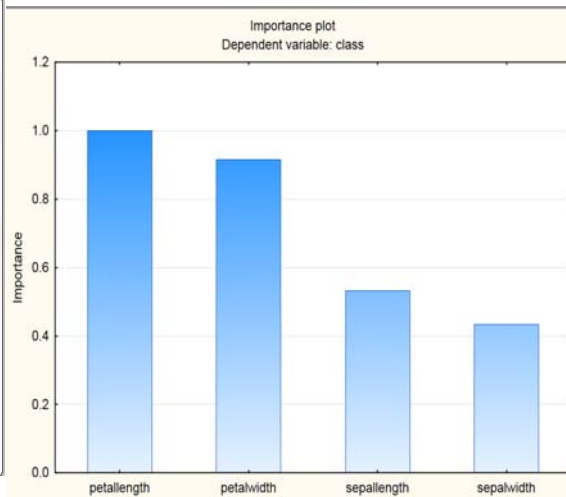
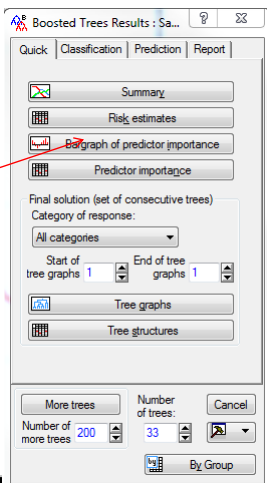


The University of Iowa

Intelligent Systems Laboratory

Boosting tree Classification

- See the results and predictor's importance

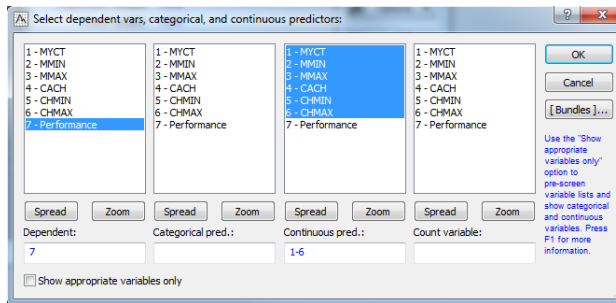
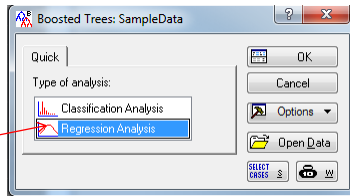


The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- CPU data set



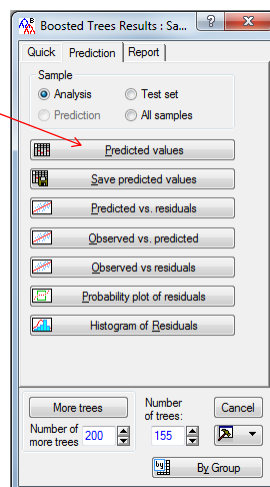
The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- See the results and predictor's importance

Predicted values



Observed value	Predicted value	Residual value
253.000	241.190	11.8101
253.000	241.190	11.8101
253.000	241.190	11.8101
290.000	264.682	25.3185
381.000	365.383	15.6168
381.000	365.383	15.6168
749.000	783.487	-34.4872
1238.000	1197.501	40.3968
24.000	26.698	-2.6982
70.000	71.814	-1.8135
117.000	110.953	6.0470
64.000	65.646	-1.5665
39.000	38.649	0.3508
40.000	37.470	2.5298
45.000	38.649	6.3508
21.000	26.698	-5.6982
28.000	31.441	-3.4409
22.000	26.698	-4.6982
28.000	27.867	0.1331
102.000	122.184	-20.1838
102.000	122.184	-20.1838
74.000	63.174	10.8261
74.000	63.174	10.8261
23.000	26.698	-3.6982
38.000	27.867	10.1331
41.000	37.246	3.7536
74.000	68.954	5.0462

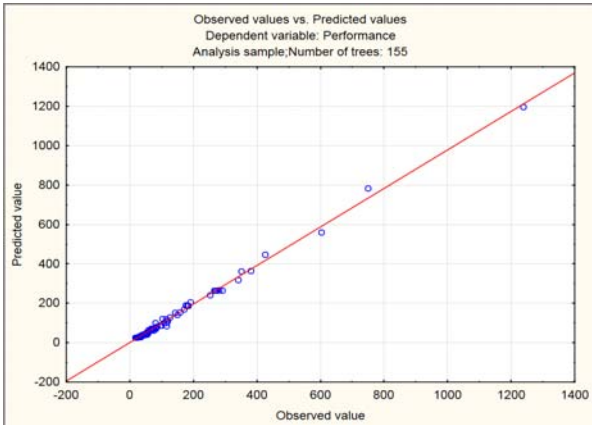
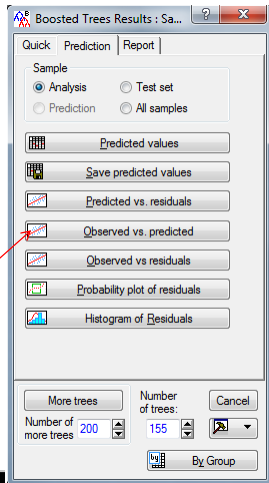


The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- See the results of Observed values vs. Predicted values

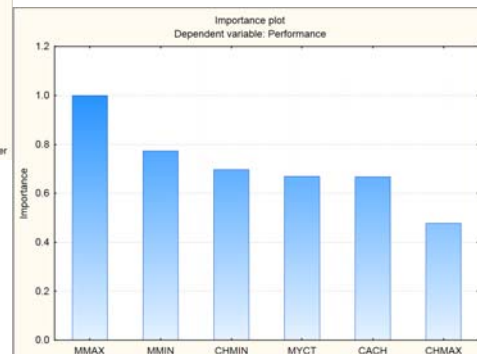
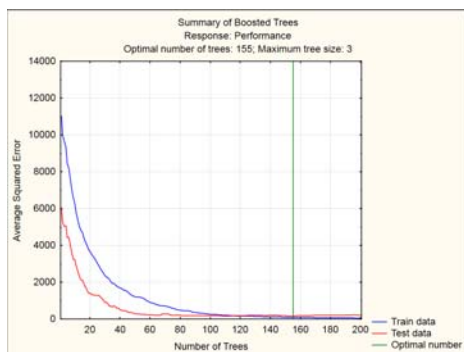


The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- See the results and predictor's importance



The University of Iowa

Neural Networks Classification

- In “Data Mining” menu and find the “Automated Neural Networks”

The screenshot shows the 'Data Mining' menu with the following options: Neural Networks, Machine Learning, GAM, Learning, IC Analysis, Optimal Binning, Cluster..., and Clustering/Grouping. A red arrow points to 'Optimal Binning'.

	1	2	3	4	5	
	sepalwidth	sepalwidth	petalwidth	petalwidth	class	
1	5.1	3.5	1.4	0.2	Iris-setosa	
2	4.9	3	1.4	0.2	Iris-setosa	
3	4.7	3.2	1.3	0.2	Iris-setosa	
4	4.6	3.1	1.5	0.2	Iris-setosa	
5	5	3.6	1.4	0.2	Iris-setosa	
6	5.4	3.9	1.7	0.4	Iris-setosa	
7	4.6	3.4	1.4	0.3	Iris-setosa	
8	5	3.4	1.5	0.2	Iris-setosa	
9	4.4	2.9	1.4	0.2	Iris-setosa	



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- Choose “Classification”, then select variables

The screenshots show the SANN software interface. The first dialog is 'New analysis/Deployment' with 'Classification' selected. The second dialog is 'Select variables for analysis' with 'class' selected as the categorical target and '1-4' as continuous inputs. The third dialog is 'Data selection: SampleData' with 'Automated network search (ANS)' selected.

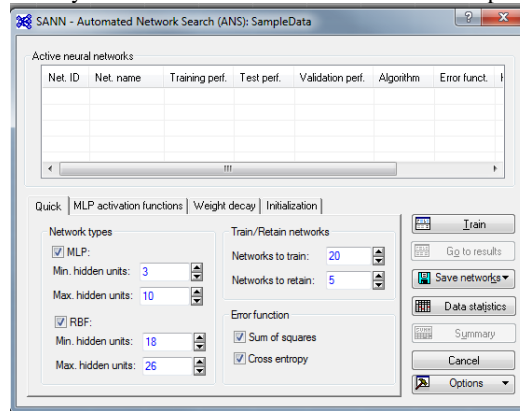


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- Statistica will try a set of different neural networks and keep the best ones

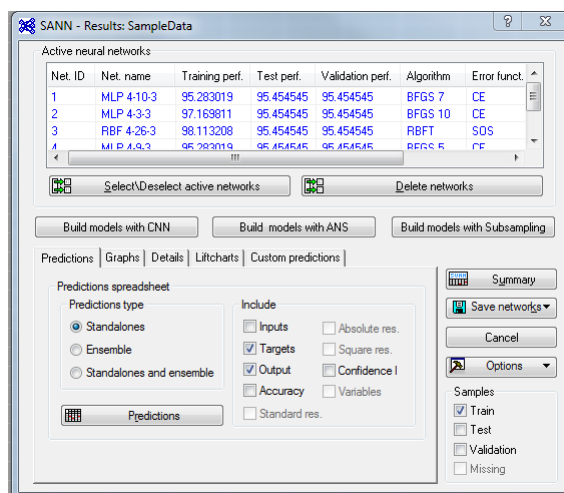


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results---Predictions

The screenshot shows the 'SANN - Results: SampleData' window. At the top, there is a table of 'Active neural networks' with columns for Net. ID, Net. name, Training perf., Test perf., Validation perf., Algorithm, and Error funct. Below this table are buttons for 'Select/Deselect active networks' and 'Delete networks'. Further down, there are buttons for 'Build models with CNN', 'Build models with ANS', and 'Build models with Subsampling'. The main area is titled 'Predictions | Graphs | Details | Liftcharts | Custom predictions |'. Under 'Predictions spreadsheet', there are radio buttons for 'Standalone', 'Ensemble', and 'Standalone and ensemble'. To the right, there are checkboxes for 'Include' options: Inputs, Targets, Output, Accuracy, Standard res., Absolute res., Square res., Confidence I, and Variables. A red arrow points to the 'Predictions' button. On the right side, there are buttons for 'Summary', 'Save networks', 'Options', and 'Cancel'.



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results---Predictions

Predictions spreadsheet for class (SampleData)						
Samples: Train						
Case name	class Target	class - Output 1_MLP 4-10-3	class - Output 2_MLP 4-3-3	class - Output 3_RBF 4-26-3	class - Output 4_MLP 4-9-3	class - Output 5_MLP 4-3-3
1	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
2	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
3	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
4	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
5	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
7	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
8	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
9	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
11	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
12	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
13	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
15	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
16	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
17	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
18	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
20	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
22	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
24	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
25	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
26	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
30	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
31	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
32	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
33	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
34	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
36	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
37	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
38	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
39	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
40	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
41	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
42	Iris-setosa	Iris-setosa	Iris-setosa	Iris-virginica	Iris-setosa	Iris-setosa
44	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
45	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
46	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
47	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
48	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa
51	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor
53	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor
55	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor
56	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor
57	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor
58	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor	Iris-versicolor



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results---Confusion matrix

SANN - Results: SampleData

Net. ID	Net. name	Training perf.	Test perf.	Validation perf.	Algorithms	Error func.
1	MLP 4-10-3	95.283019	95.454545	95.454545	BFGS 7	CE
2	MLP 4-3-3	97.168811	95.454545	95.454545	BFGS 10	CE
3	RBF 4-26-3	98.112300	95.454545	95.454545	RBF1	SOS
4	MLP 4-9-3	95.303016	95.454545	95.454545	RBF1	FC

Build models with CNN | Build models with ANS | Build models with Subsampling

Predictions | Graphs | Details | Litcharts | Custom predictions

Summary

Weights | Predictions statistics

Correlation coefficients | Data statistics

Confusion matrix | Global sensitivity analysis

Confidence levels | Local sensitivity analysis

Predicted category	class (Confusion matrix) (SampleData)		
	class-Iris-setosa	class-Iris-versicolor	class-Iris-virginica
1.MLP 4-10-3-Iris-setosa	37	0	0
1.MLP 4-10-3-Iris-versicolor	0	33	4
1.MLP 4-10-3-Iris-virginica	0	1	31
2.MLP 4-3-3-Iris-setosa	37	0	0
2.MLP 4-3-3-Iris-versicolor	0	34	3
2.MLP 4-3-3-Iris-virginica	0	0	32
3.RBF 4-26-3-Iris-setosa	36	0	0
3.RBF 4-26-3-Iris-versicolor	0	33	0
3.RBF 4-26-3-Iris-virginica	1	1	35
4.MLP 4-9-3-Iris-setosa	37	0	0
4.MLP 4-9-3-Iris-versicolor	0	32	3
4.MLP 4-9-3-Iris-virginica	0	2	32
5.MLP 4-3-3-Iris-setosa	37	0	0
5.MLP 4-3-3-Iris-versicolor	0	32	0
5.MLP 4-3-3-Iris-virginica	0	2	35



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Regression

- CPU data set

	1	2	3	4	5	6	7
	MYCT	MMIN	MMAx	CACH	CHMIN	CHMAX	Performance
1	125	256	6000	256	16	128	199
2	29	8000	32000	32	8	32	253
3	29	8000	32000	32	8	32	253
4	29	8000	32000	32	8	32	253
5	29	8000	16000	32	8	16	132
6	26	8000	32000	64	8	32	290
7	23	16000	32000	64	16	32	381
8	23	16000	32000	64	16	32	381
9	23	16000	64000	64	16	32	749
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28	320	512	5000	4	1	5	28
29	320	256	5000	4	1	6	27
30	25	1310	2620	131	12	24	102
31	25	1310	2620	131	12	24	102
32	25	1310	2620	131	12	24	102
33	50	2620	10480	30	12	24	74

SANN - New Analysis/Deployment: SampleData

New analysis/Deployment

Deployment

Deploy models from previous analyses

New analysis

File name | Net. ID | Net. name | Hidden act. | Output act.

PHML file list

Regression

Classification

Time series (regression)

Time series (classification)

Cluster analysis

Select an analysis type from the list above to start a new analysis. To deploy models from previous analyses, use the deployment option.

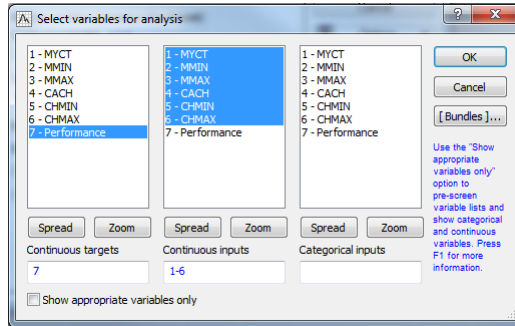


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Regression

- CPU data set, select variables

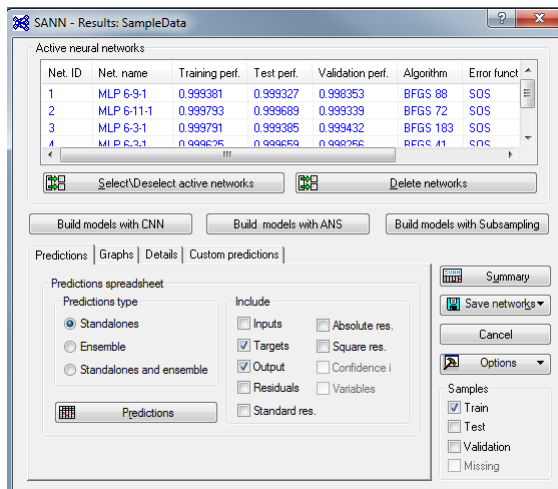


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Regression

- Training and results



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Regression

- Predictions

Predictions spreadsheet for Performance (SampleData)

Case name	Performance Target	Performance - Output 1 MLP 6-9-1	Performance - Output 2 MLP 6-11-1	Performance - Output 3 MLP 6-3-1	Performance - Output 4 MLP 6-3-1	Performance - Output 5 MLP 6-7-1
1	195 000	195 825	199 225	201 352	201 262	202 672
2	283 000	244 271	246 196	246 521	240 423	245 611
3	253 000	244 271	246 196	246 521	240 423	245 611
8	381 000	368 668	373 375	375 731	363 537	374 480
9	749 000	747 172	750 181	751 373	753 234	753 595
10	1238 000	1236 946	1238 243	1237 226	1237 902	1236 850
13	70 000	70 524	69 341	71 225	72 291	72 390
14	117 000	120 840	116 328	117 726	119 930	119 054
15	15 000	20 605	16 708	15 167	14 425	15 214
16	54 000	59 500	56 529	65 026	65 120	65 102
17	23 000	28 280	22 444	23 369	22 254	22 275
18	29 000	35 338	29 838	30 903	29 252	29 754
19	22 000	28 436	21 596	22 033	21 054	20 425
20	124 000	129 883	128 786	127 641	130 862	129 085
22	39 000	40 302	40 482	40 196	41 839	39 075
23	40 000	46 341	43 461	41 761	42 620	41 359
26	21 000	22 907	21 230	20 700	20 521	20 732
27	28 000	26 271	29 305	28 403	27 768	29 186
28	22 000	23 541	22 965	22 416	21 936	22 985
29	26 000	25 981	26 120	27 731	27 528	28 209
30	27 000	25 788	27 712	27 293	26 762	28 013
31	102 000	95 155	101 303	93 937	93 049	92 748
32	102 000	95 155	101 303	93 937	93 049	92 748
33	74 000	75 389	74 475	76 151	77 039	76 827
34	74 000	75 389	74 475	76 151	77 039	76 827
35	138 000	142 035	146 438	143 815	146 681	145 540
37	23 000	28 440	21 298	19 865	20 757	20 240
38	29 000	32 558	27 784	27 214	27 905	27 875
39	44 000	43 655	43 032	43 523	44 039	44 546
40	30 000	33 606	28 144	28 872	28 383	28 540
41	41 000	42 377	40 754	41 988	40 614	41 921
42	74 000	73 071	75 984	76 688	75 463	76 951
43	74 000	73 345	76 196	76 922	75 607	77 160
44	74 000	73 345	76 196	76 922	75 607	77 160
45	54 000	51 209	55 317	55 485	54 962	55 842
47	18 000	18 372	19 874	20 779	20 827	21 684
48	28 000	19 813	26 514	26 791	27 438	26 490
50	38 000	34 769	38 133	37 656	37 582	38 515



The University of Iowa

Intelligent Systems Laboratory

Neural Networks Regression

- Some statistics about the predictions

Data statistics (SampleData)

Samples	MYCT Input	MMIN Input	MMAX Input	CACH Input	CHMIN Input	CHMAX Input	Performance Target
Minimum (Train)	17 000	64 00	64 00	0 0000	1 00000	1 0000	15 000
Maximum (Train)	1100 000	32000 00	64000 00	256 0000	32 00000	128 0000	1238 000
Mean (Train)	192 151	2963 02	11747 58	25 3237	4 52516	17 8201	96 568
Standard deviation (Train)	229 786	4174 75	11421 62	41 9709	5 40576	21 5814	145 167
Minimum (Test)	25 000	512 00	1000 00	0 0000	0 00000	0 0000	18 000
Maximum (Test)	1500 000	16000 00	32000 00	160 0000	16 00000	38 00000	426 000
Mean (Test)	190 172	2993 66	9706 90	26 3793	3 24138	10 8621	76 172
Standard deviation (Test)	303 963	3336 71	7155 81	38 4136	3 26950	8 6550	86 729
Minimum (Validation)	23 000	96 00	512 00	0 0000	0 00000	0 0000	15 000
Maximum (Validation)	1500 000	16000 00	32000 00	128 0000	52 00000	104 0000	381 000
Mean (Validation)	284 069	2540 41	11689 03	20 1034	6 00000	18 2069	88 793
Standard deviation (Validation)	184 035	7624 24	17931 45	59 7860	6 97444	29 2859	275 283
Minimum (Missing)							
Maximum (Missing)							
Mean (Missing)							
Std (Missing)							
Minimum (Overall)	17 000	64 00	64 00	0 0000	0 00000	0 0000	15 000
Maximum (Overall)	1500 000	32000 00	64000 00	256 0000	52 00000	128 0000	1238 000
Mean (Overall)	205 391	2905 32	11438 56	24 7107	4 55330	16 8528	92 421
Standard deviation (Overall)	265 596	3934 15	10680 22	40 1302	6 05615	20 3599	132 087



The University of Iowa

Intelligent Systems Laboratory