Data Mining: STATISTICA

The University of Iowa

Intelligent Systems Laboratory

Outline

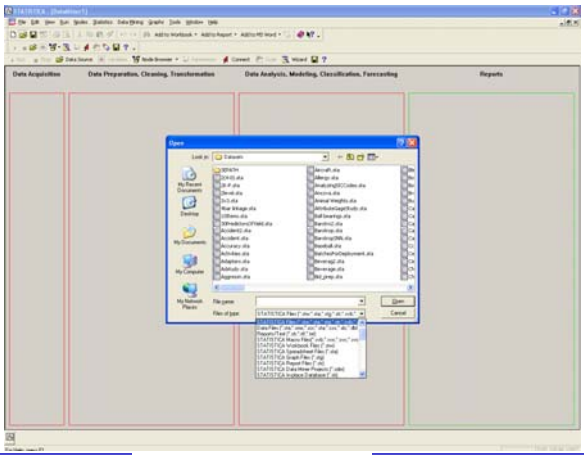
- Prepare the data
- Classification and regression

The University of Iowa

Intelligent Systems Laboratory

Prepare the Data

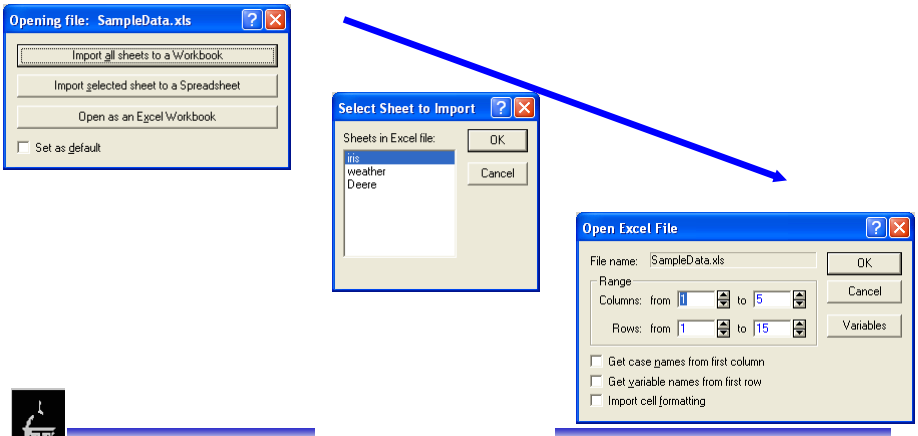
- Statistica can read from Excel, .txt and many other types of files
- Compared with WEKA, Statistica is much easier in terms of data preparing



The University of Iowa Intelligent Systems Laboratory

Open an Excel File

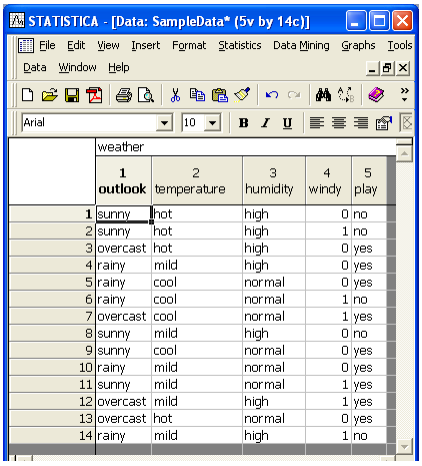
- Click the “Import selected sheet to Spreadsheet”
- Select the desired Excel sheet where your data is stored
- Get variable names from the first row



The University of Iowa Intelligent Systems Laboratory

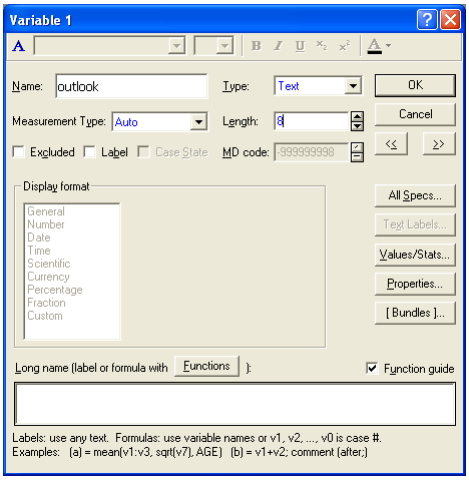
Open an Excel File

- Change variable type



Open an Excel File

- Change variable type

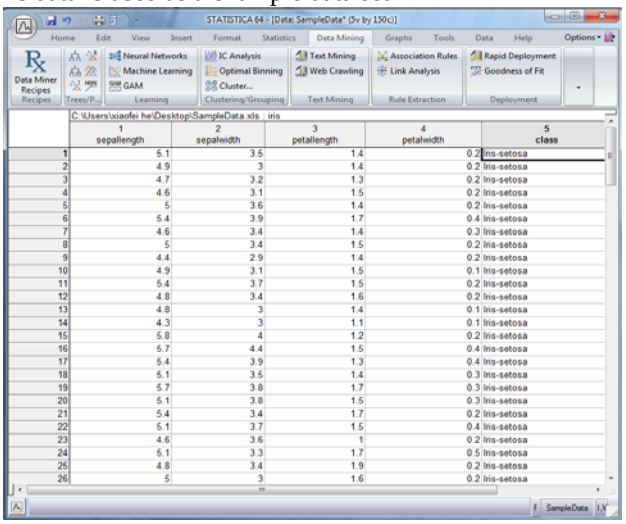


Classification and Regression

- C&RT
- Boosting tree
- Neural Networks

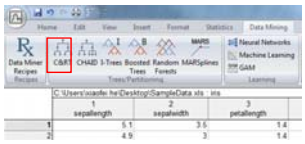
C&RT Classification

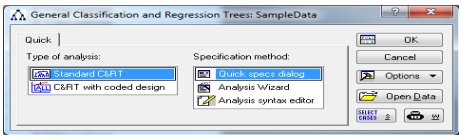
- Iris data is used as a example data set

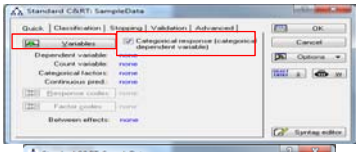


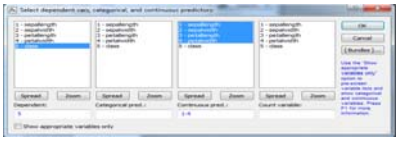
C&RT Classification

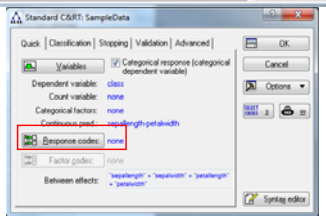
- Click “Data Mining” menu and find the “Interactive Trees”

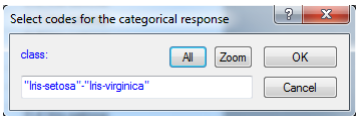










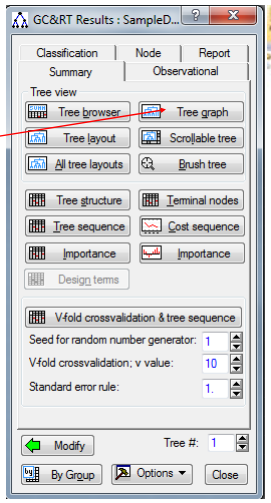


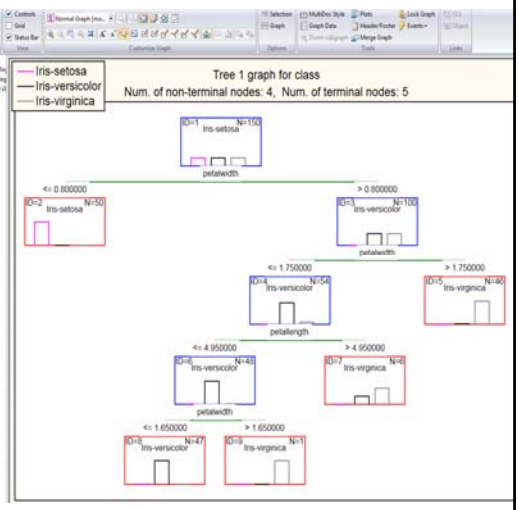
The University of Iowa

Intelligent Systems Laboratory

C&RT Classification

- View the final tree and understand the results



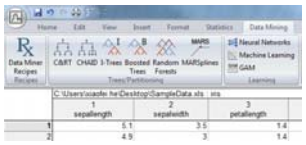


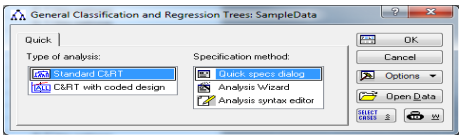
The University of Iowa

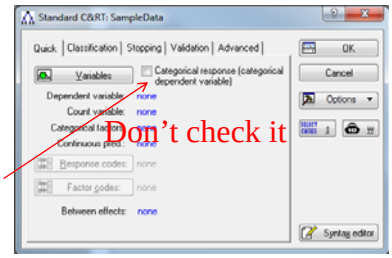
Intelligent Systems Laboratory

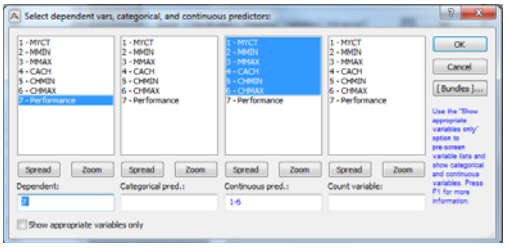
C&RT---Regression

- Use the CPU data set and select the regression analysis







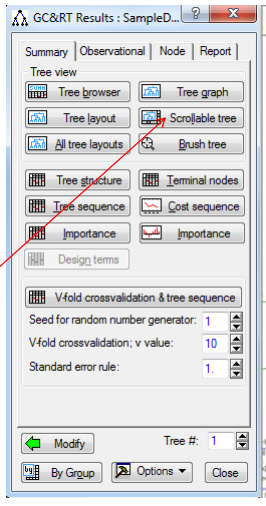


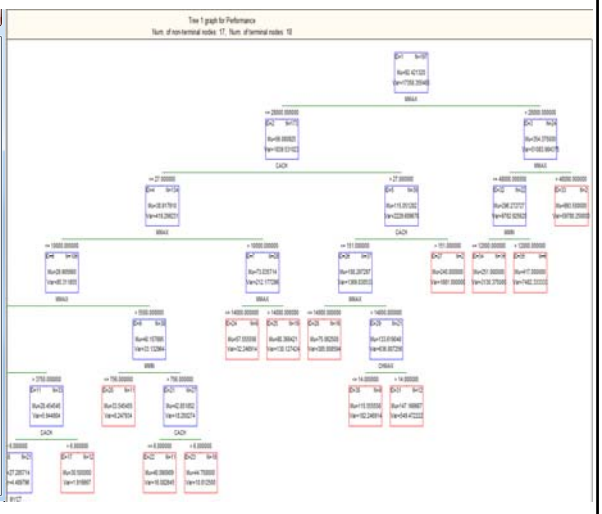
The University of Iowa

Intelligent Systems Laboratory

C&RT---Regression

- Regression tree structure

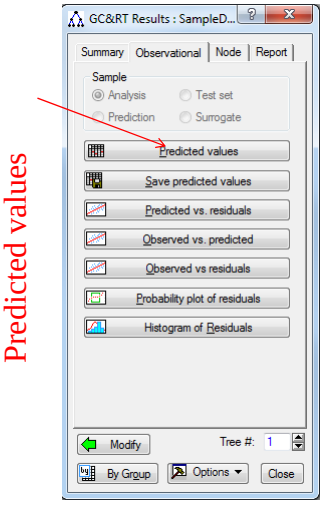




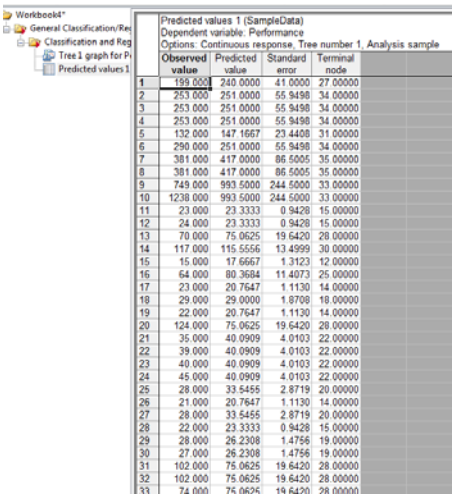
The University of Iowa

Intelligent Systems Laboratory

C&RT---Regression

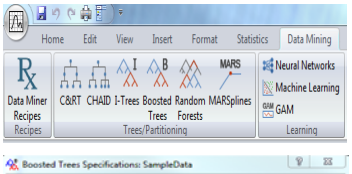


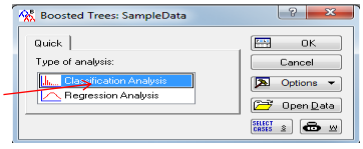
Predicted values

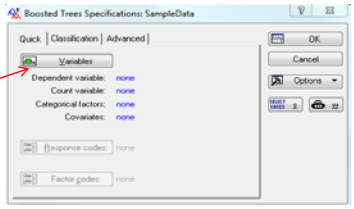


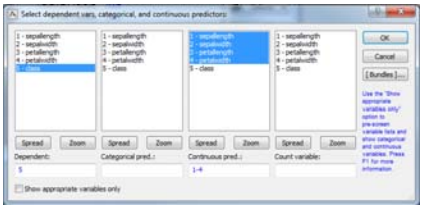
Boosting tree Classification

- In “Data Mining” menu and find the “Boosted Trees”



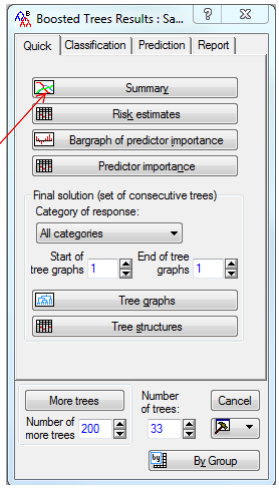


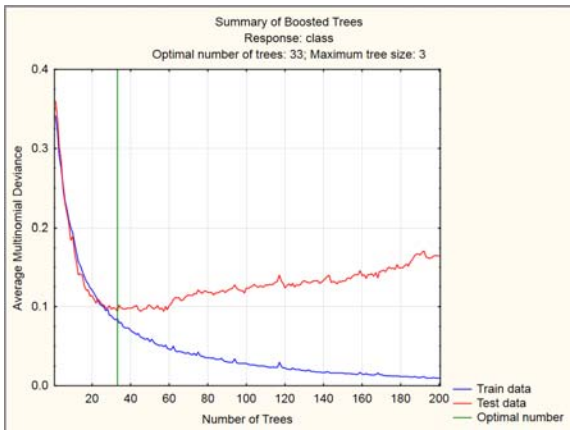




Boosting tree Classification

- See the results and predictor's importance



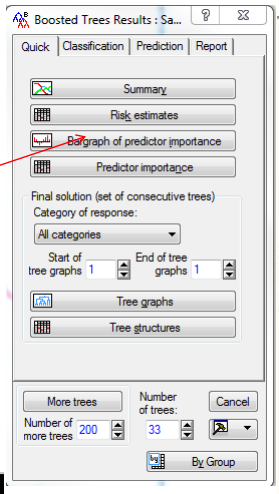


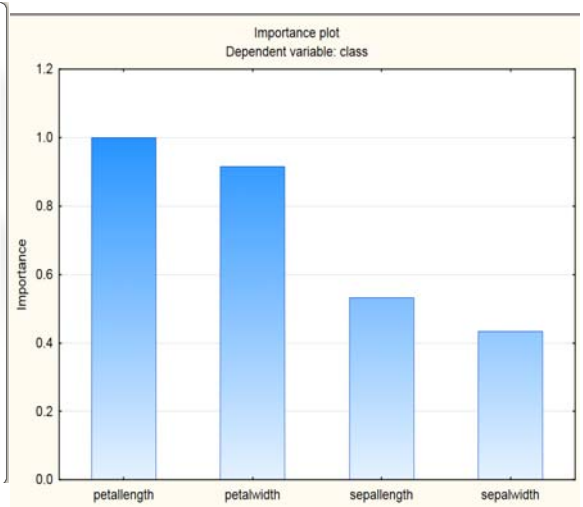
The University of Iowa

Intelligent Systems Laboratory

Boosting tree Classification

- See the results and predictor's importance



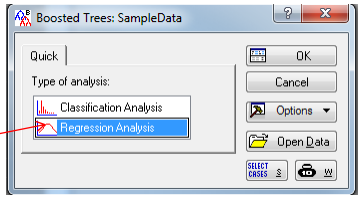


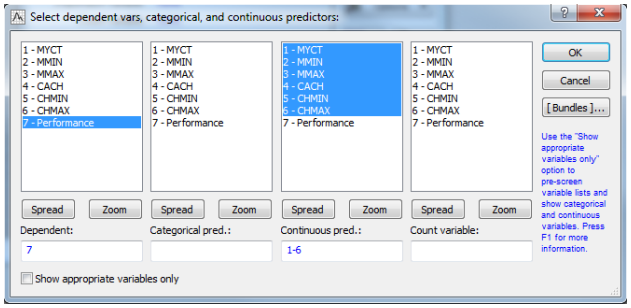
The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- CPU data set



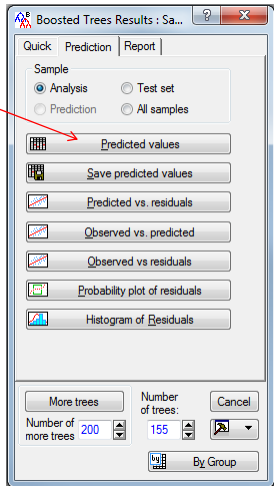


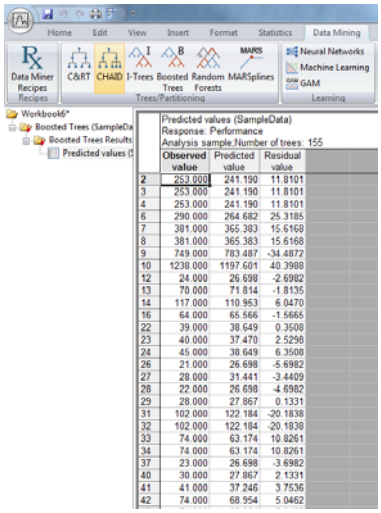
The University of Iowa — Intelligent Systems Laboratory

Boosting tree Regression

- See the results and predictor's importance

Predicted values

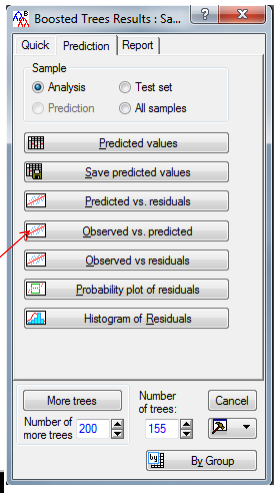


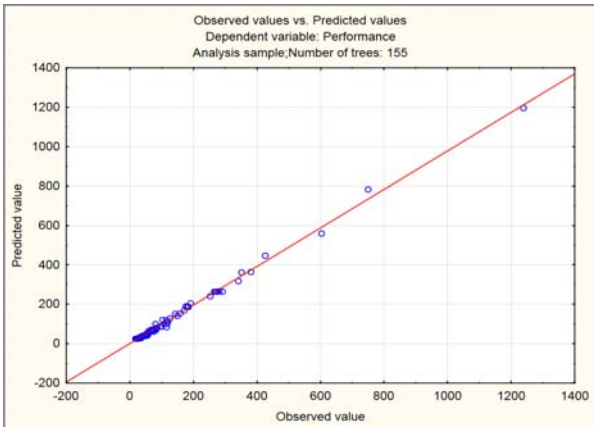


The University of Iowa — Intelligent Systems Laboratory

Boosting tree Regression

- See the results of Observed values vs. Predicted values



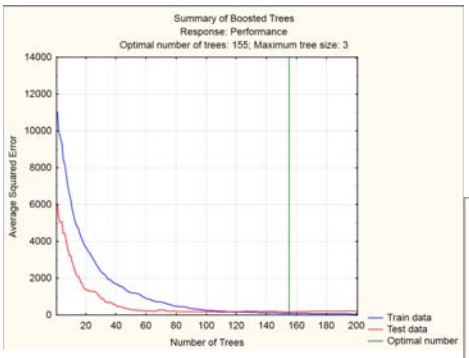


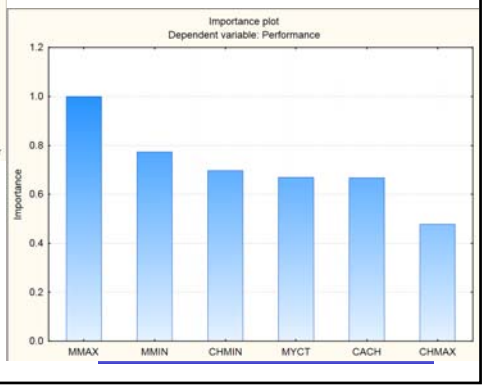
The University of Iowa

Intelligent Systems Laboratory

Boosting tree Regression

- See the results and predictor's importance

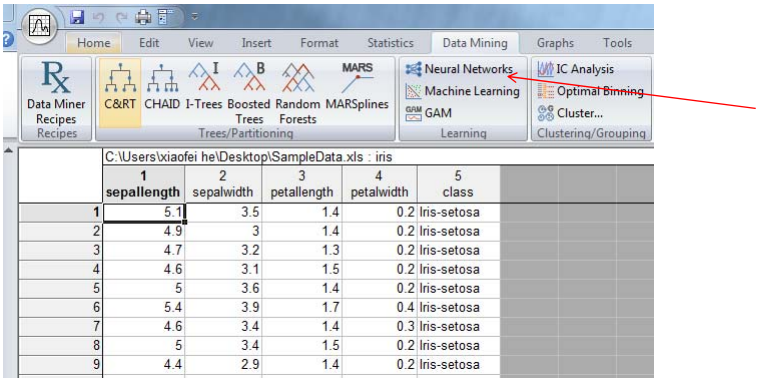




The University of Iowa

Neural Networks Classification

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

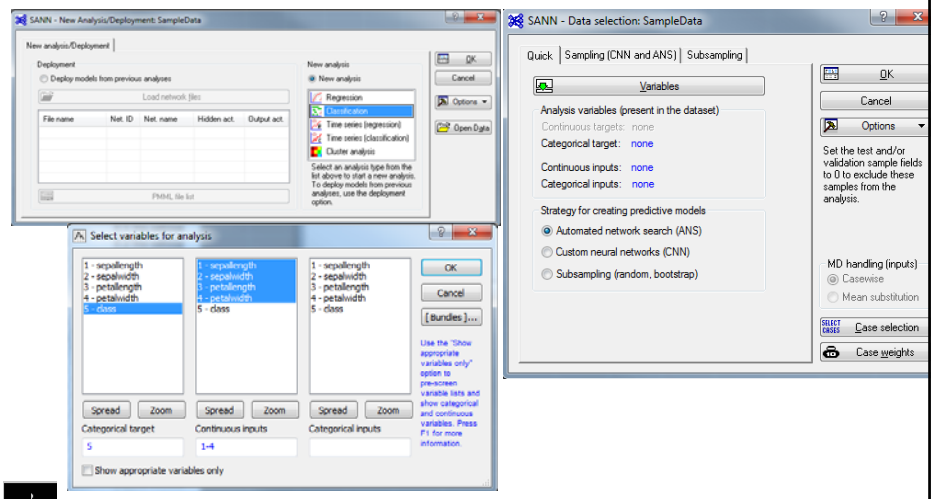


	1 sepallength	2 sepalwidth	3 petallength	4 petalwidth	5 class
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4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

The University of Iowa Intelligent Systems Laboratory

Neural Networks Classification

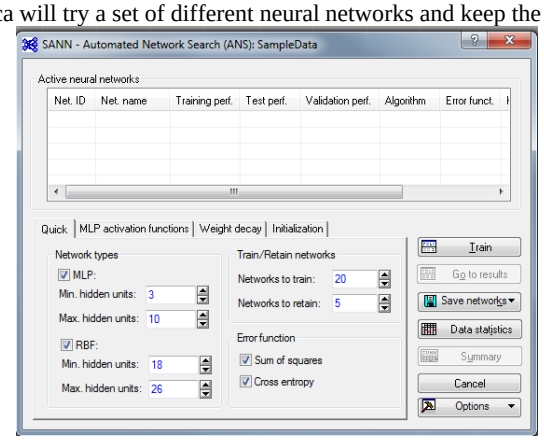
- Choose “Classification”, then select variables



The University of Iowa Intelligent Systems Laboratory

Neural Networks Classification

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

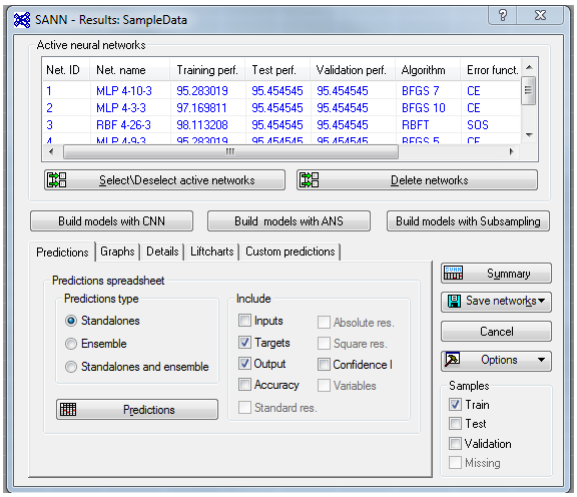


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results

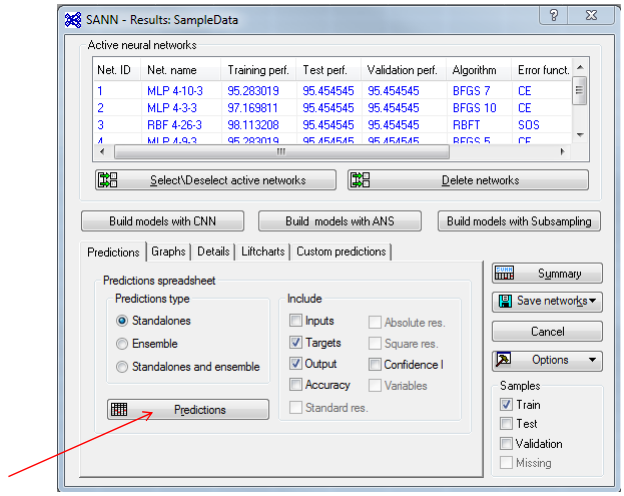


The University of Iowa

Intelligent Systems Laboratory

Neural Networks Classification

- See the classification results---Predictions



Neural Networks Classification

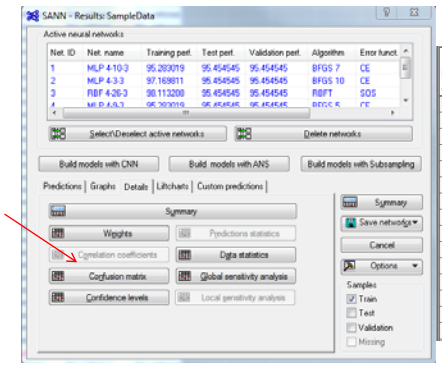
- See the classification results---Predictions

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-virginica	Iris-versicolor	Iris-versicolor	Iris-virginica	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

Neural Networks Classification

- See the classification results---Confusion matrix



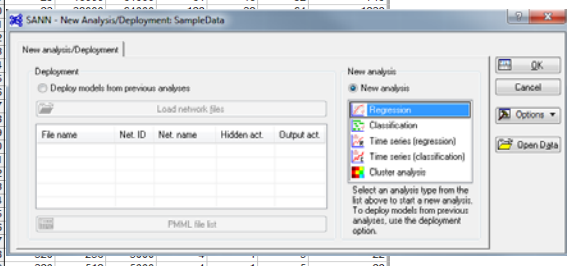
Predicted category	class (Confusion matrix) (SampleData) Samples: Train		
	class-Iris-setosa	class-Iris-versicolor	class-Iris-virginica
1.MLP 4-10-3-Iris-setosa	37	0	0
1.MLP 4-10-3-Iris-versicolor	0	33	4
1.MLP 4-10-3-Iris-virginica	0	1	31
2.MLP 4-3-3-Iris-setosa	37	0	0
2.MLP 4-3-3-Iris-versicolor	0	34	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

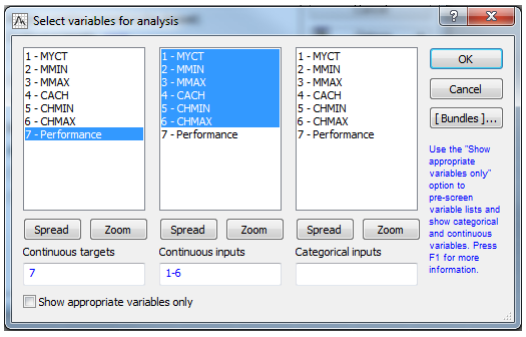
	1 MYCT	2 MMIN	3 MMAX	4 CACH	5 CHMIN	6 CHMAX	7 Performance
1	125	256	6000	256	16	128	199
2	29	8000	32000	32	8	32	253
3	29	8000	32000	32	8	32	253
4	29	8000	32000	32	8	32	253
5	29	8000	16000	32	8	16	132
6	26	8000	32000	64	8	32	290
7	23	16000	32000	64	16	32	381
8	23	16000	32000	64	16	32	381
9	23	16000	64000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



The University of Iowa Intelligent Systems Laboratory

Neural Networks Regression

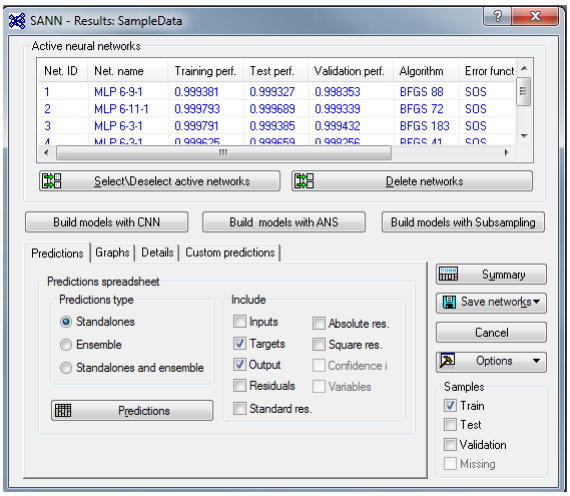
- CPU data set, select variables



The University of Iowa Intelligent Systems Laboratory

Neural Networks Regression

- Training and results



The University of Iowa Intelligent Systems Laboratory

Neural Networks Regression

- Predictions

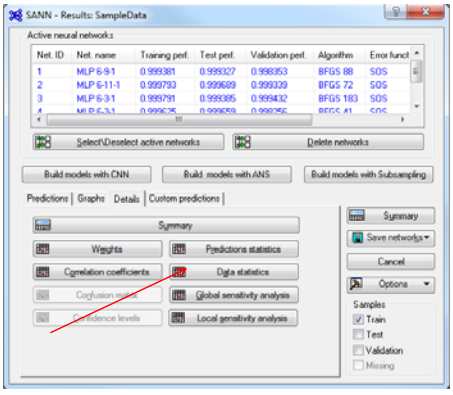
Predictions spreadsheet for Performance (SampleData)
Samples: Train

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	199.000	195.829	199.229	201.352	201.282	202.672
2	253.000	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	753.505
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	64.000	59.500	66.929	65.026	65.120	65.702
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	41.639	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	28.000	25.981	28.120	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	27.314	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	37.606	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	4.52518	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	38.0000	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