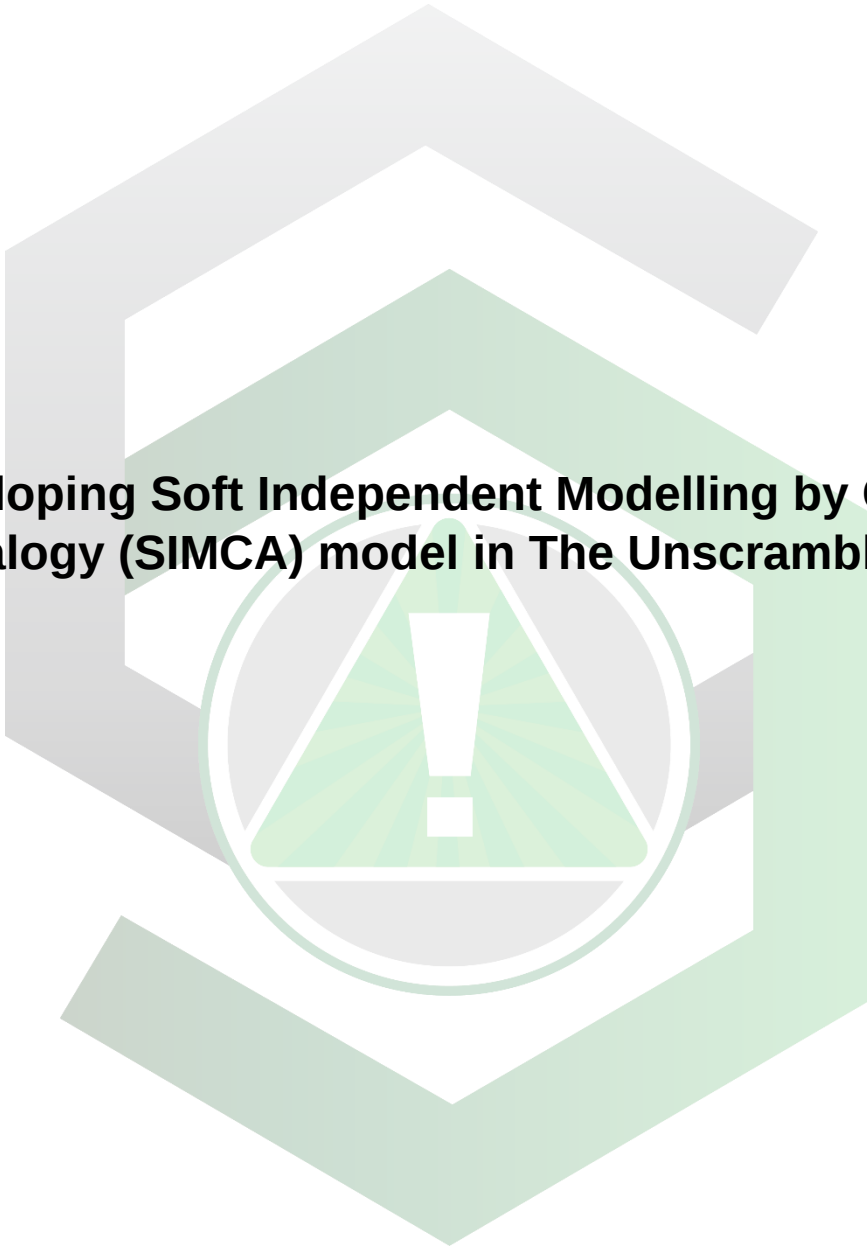


The background features a large, stylized hexagonal shape composed of concentric, slightly offset layers in light gray and light green. In the center of this graphic is a circular warning sign. The sign has a green border and a green background with a white exclamation mark. The text is centered over the warning sign.

Developing Soft Independent Modelling by Class Analogy (SIMCA) model in The UnscramblerX

Created by: Edwin Caballero-Agosto

Manager: Samuel Hernandez-Rivera

SOP-01	Edwin Caballero-Agosto		University of Puerto Rico at Mayagüez
Effectivity: Apr/08/2022	Developing Soft Independent Modelling by Class Analogy model in The UnscramblerX		Revised by:
Revised:			Approved by:

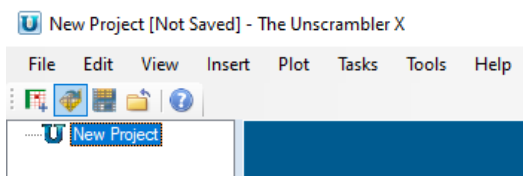
PREPARING DATA FOR ANALYSIS

This SOP uses the “Tutorial H” data set from The UnscramblerX program PCA tutorial.

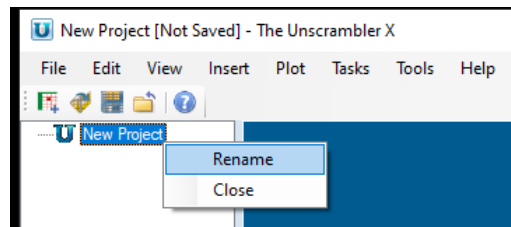
1. **Create** Excel file (.xlsx).
2. **Write** data with samples in rows and variables in columns.

	A	B	C	D	E	F
1	Sample	Class	Sepal length	Sepal width	Petal length	Petal width
2	setosa 1	setosa	5.1	3.5	1.4	0.2
3	setosa 2	setosa	4.9	3.0	1.4	0.2
4	setosa 3	setosa	4.7	3.2	1.3	0.2
5	setosa 4	setosa	4.6	3.1	1.5	0.2
6	setosa 5	setosa	5.0	3.6	1.4	0.2
7	setosa 6	setosa	5.4	3.9	1.7	0.4
8	setosa 7	setosa	4.6	3.4	1.4	0.3

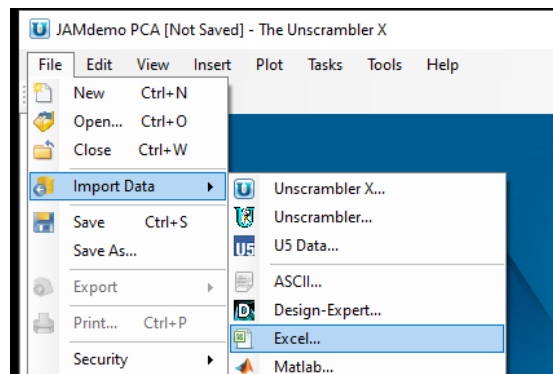
3. **Save** spreadsheet file as “SIMCA Data.xlsx” in desktop by going to File → Save As... on the Desktop area.
4. **Open** The UnscramblerX program.



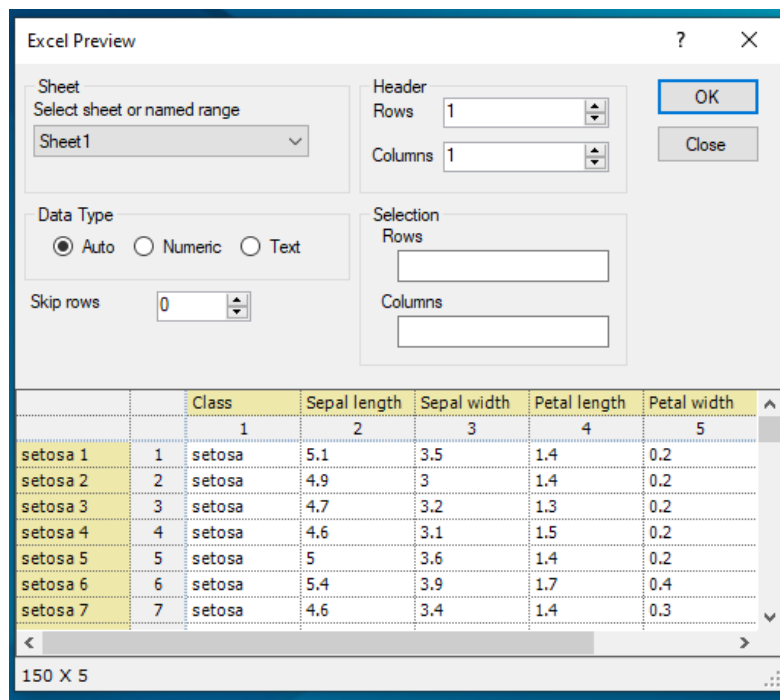
5. **Rename** project with right click → Rename to “SIMCA Development”.



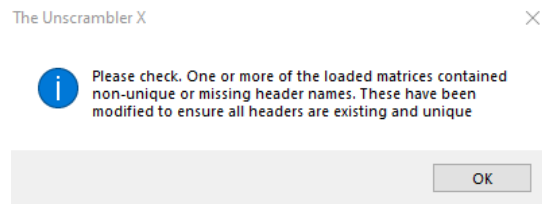
6. **Import** data with File → Import Data → Excel..., select the created file in the Desktop area, click “Open”.



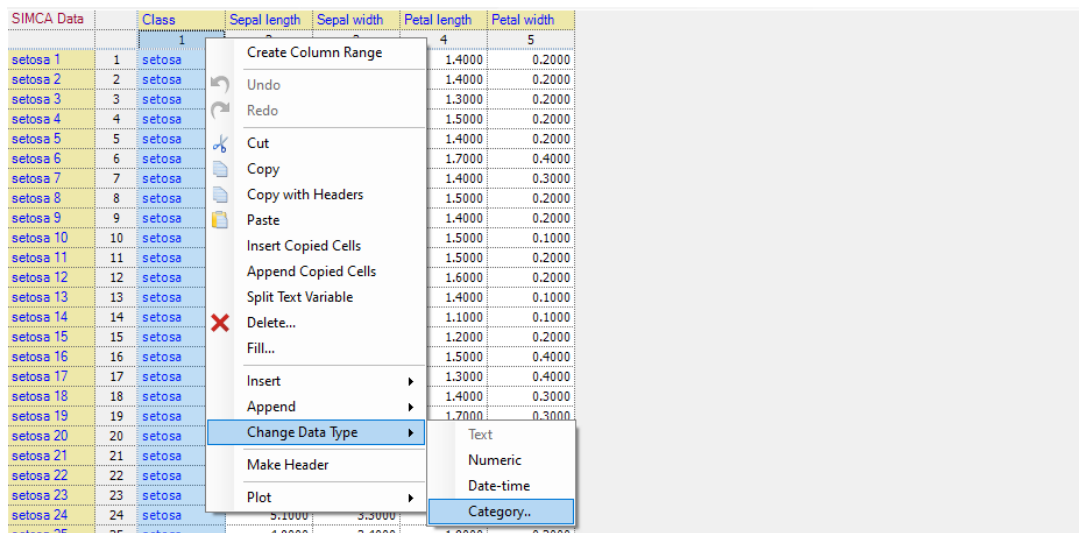
7. **Select** “Sheet 1” and press “OK” on the “Excel Preview” window.



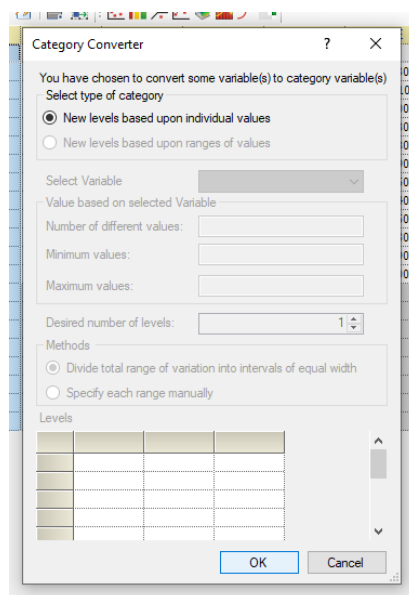
- Click “OK” on the pop-up window if it appears.



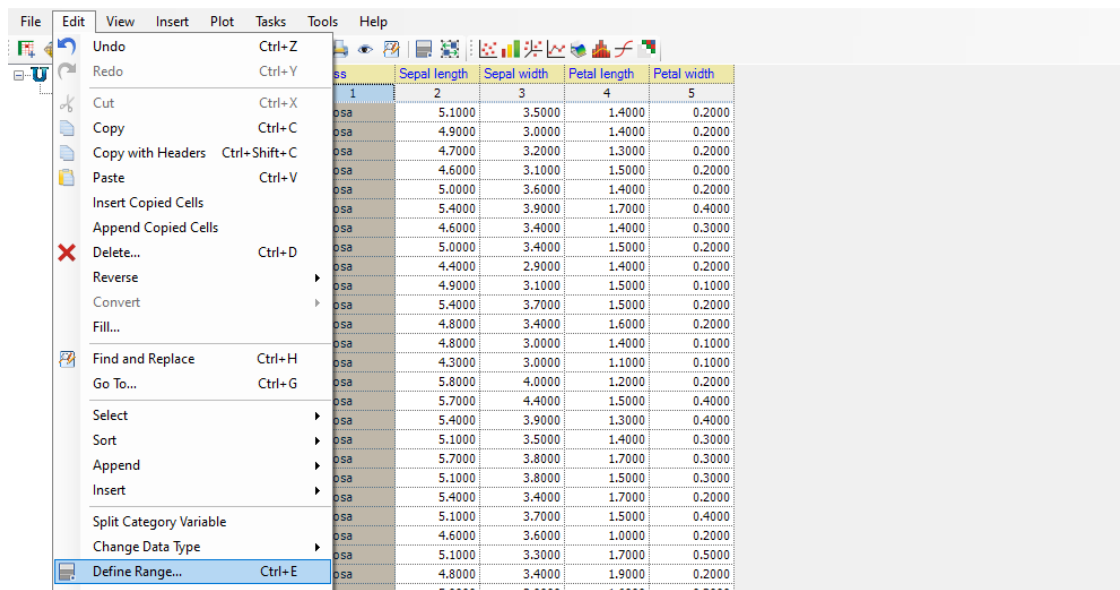
- Change the “Class” column from text to category by left click on the “1” below the “Class” text, right click → Change Data Type → Category...



- Click “OK” on the “Category Converter” window.

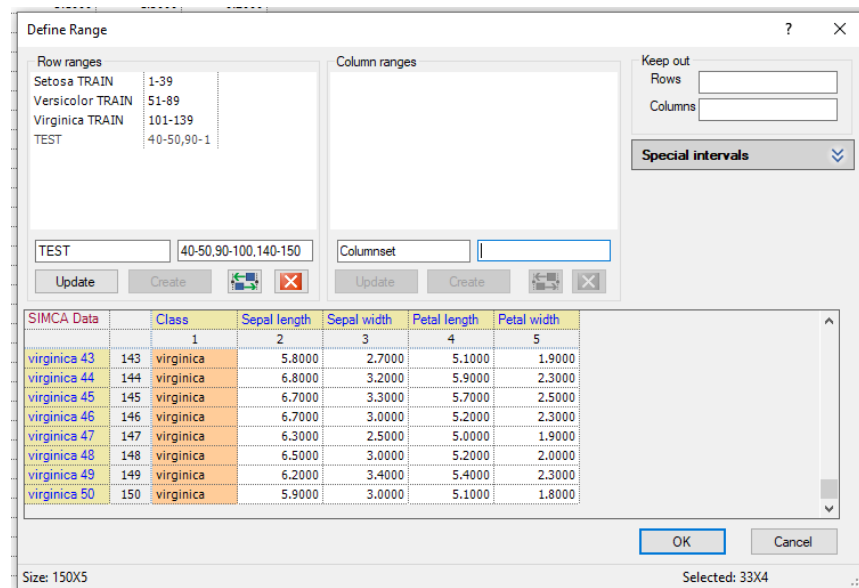


11. Create sample and variable sets with Edit → Define Range...

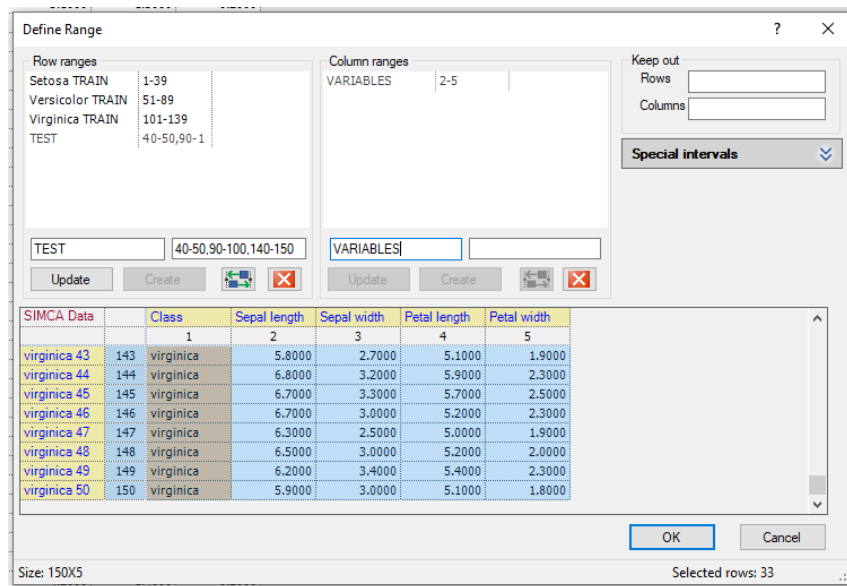


12. Create “TRAIN” sample set with the data that will be used to create the PCA models for each class and “TEST” set for the data that will be used to test the SIMCA model. Click “Create”.

In this example, there are 50 samples per class (setosa, versicolor, and virginica). Samples 1-39 were used to create the TRAIN set and samples 40-50 were used to create the TEST set.

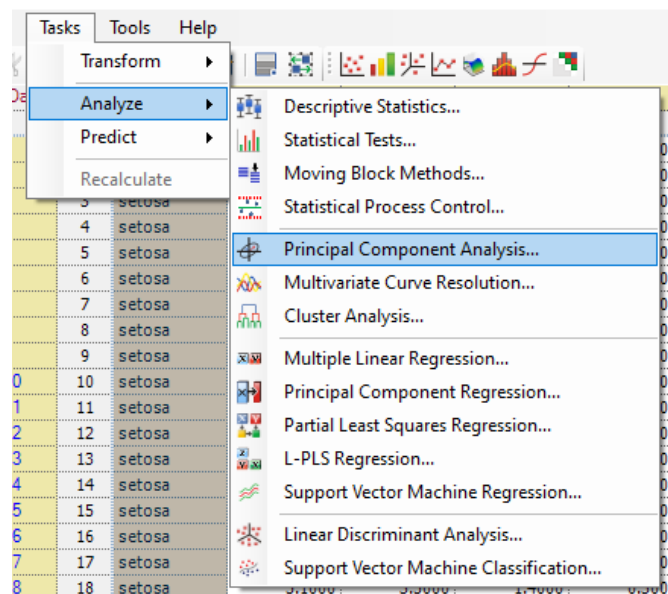


13. Create set with variables on the column ranges section.

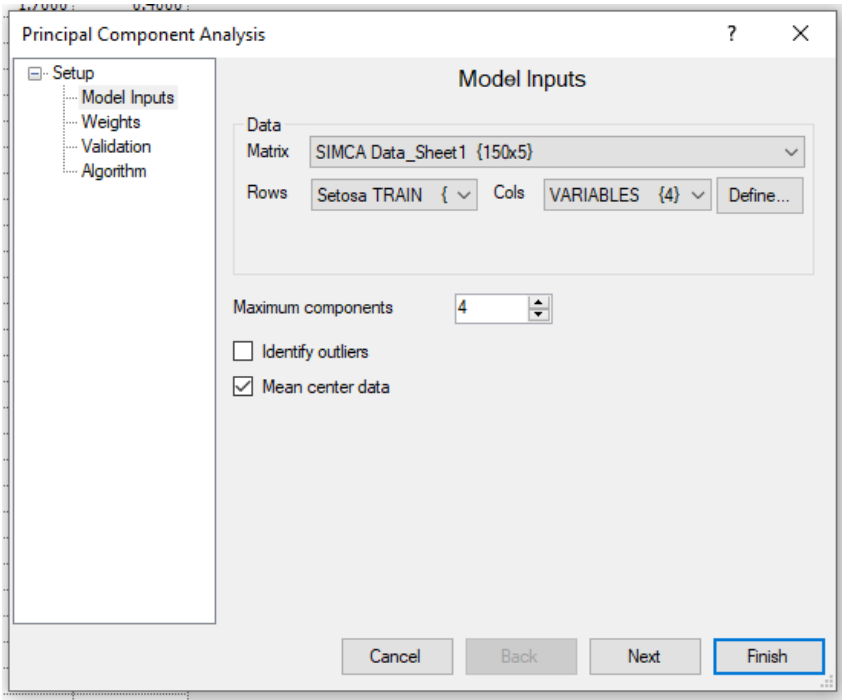


14. Click "OK" to save sets. **WARNING:** If you press "ENTER" you will lose all the sets created.

15. Develop PCA model with Tasks → Analyze → Partial Least Squares...



16. **Select** “SIMCA Data_Sheet” matrix, select “Setosa TRAIN” rows set and “VARIABLES” cols set. Choose 4 maximum components.



The dialog box is titled "Principal Component Analysis". On the left, a tree view shows "Setup" expanded, with sub-items "Model Inputs", "Weights", "Validation", and "Algorithm". The "Model Inputs" tab is active. The "Data Matrix" is set to "SIMCA Data_Sheet1 {150x5}". The "Rows" are set to "Setosa TRAIN" and the "Cols" are set to "VARIABLES {4}". The "Maximum components" is set to 4. The "Identify outliers" checkbox is unchecked, and the "Mean center data" checkbox is checked. At the bottom are "Cancel", "Back", "Next", and "Finish" buttons.

Principal Component Analysis

Model Inputs

Data Matrix: SIMCA Data_Sheet1 {150x5}

Rows: Setosa TRAIN { } Cols: VARIABLES {4} Define...

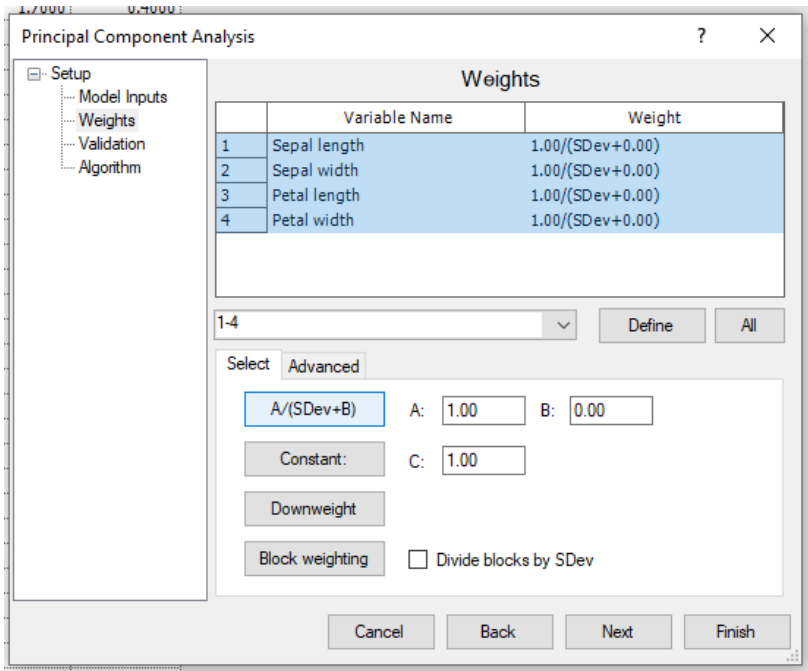
Maximum components: 4

☐ Identify outliers

☒ Mean center data

Cancel Back Next Finish

17. **Select** “All”, select “A/(SDev+B)” in the “Weights” tab. If data is composed of spectra, do not modify the “Weights” tab.



The dialog box is titled "Principal Component Analysis". On the left, the "Weights" tab is selected in the tree view. The "Weights" tab displays a table with 4 rows and 2 columns: "Variable Name" and "Weight". The rows are: 1. Sepal length, 2. Sepal width, 3. Petal length, 4. Petal width. All weights are set to "1.00/(SDev+0.00)". Below the table, a dropdown menu shows "1-4". To the right of the dropdown are "Define" and "All" buttons. Below these are "Select" and "Advanced" tabs. The "Advanced" tab is active, showing a list of weighting options: "A/(SDev+B)" (selected), "Constant:", "Downweight", and "Block weighting". To the right of these options are input fields for A, B, and C. A checkbox "Divide blocks by SDev" is also present. At the bottom are "Cancel", "Back", "Next", and "Finish" buttons.

Principal Component Analysis

Weights

	Variable Name	Weight
1	Sepal length	1.00/(SDev+0.00)
2	Sepal width	1.00/(SDev+0.00)
3	Petal length	1.00/(SDev+0.00)
4	Petal width	1.00/(SDev+0.00)

1-4 Define All

Select Advanced

A/(SDev+B) A: 1.00 B: 0.00

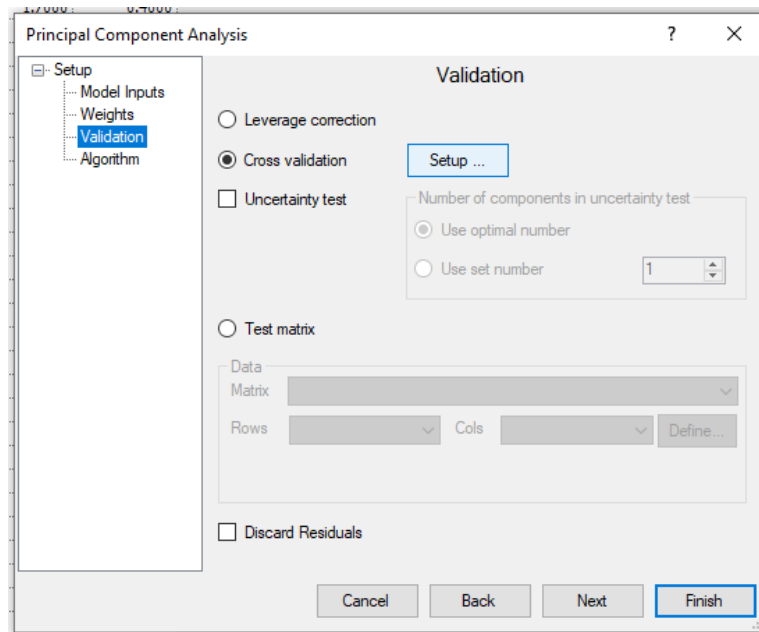
Constant: C: 1.00

Downweight

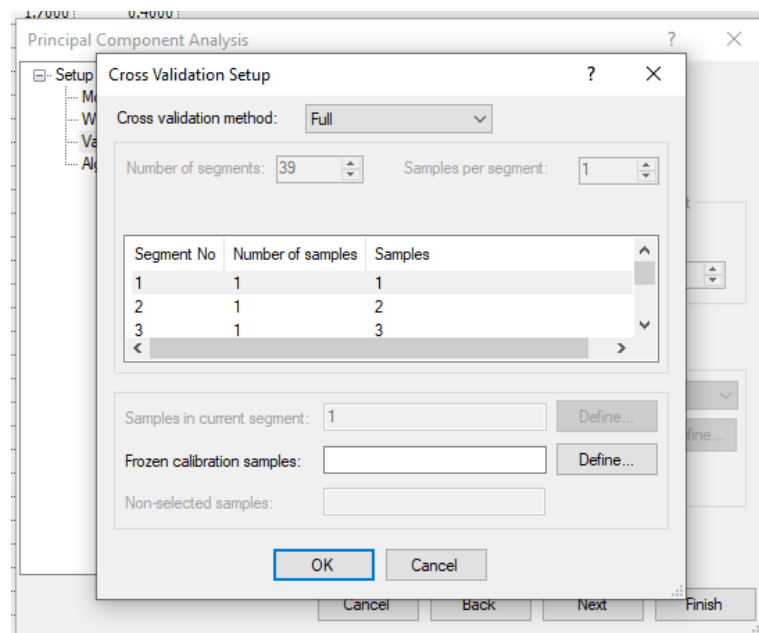
Block weighting ☐ Divide blocks by SDev

Cancel Back Next Finish

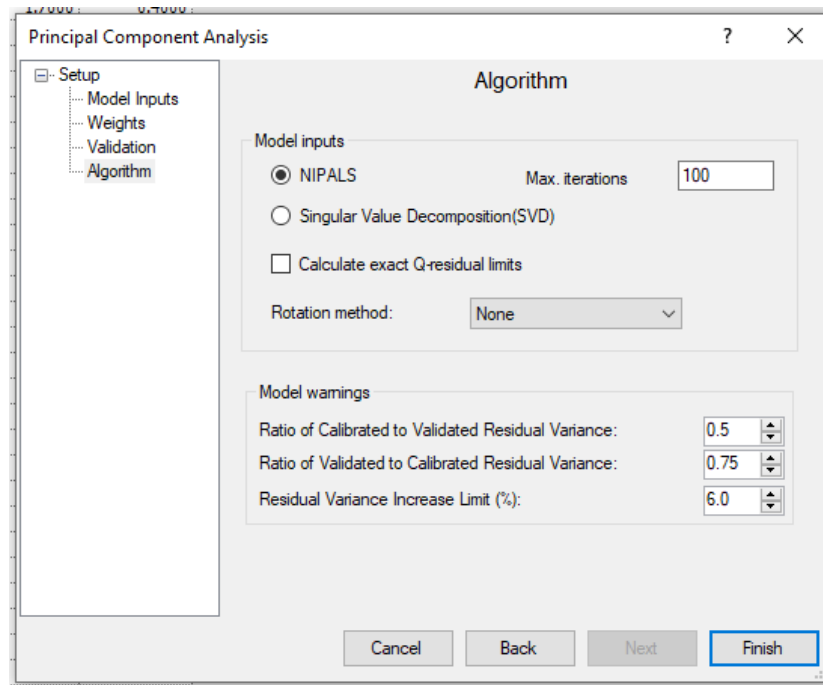
18. Select “Setup ...” button in “Validation” tab.



19. Select “Full” cross validation method in the “Cross Validation Setup” window, and press “OK”.



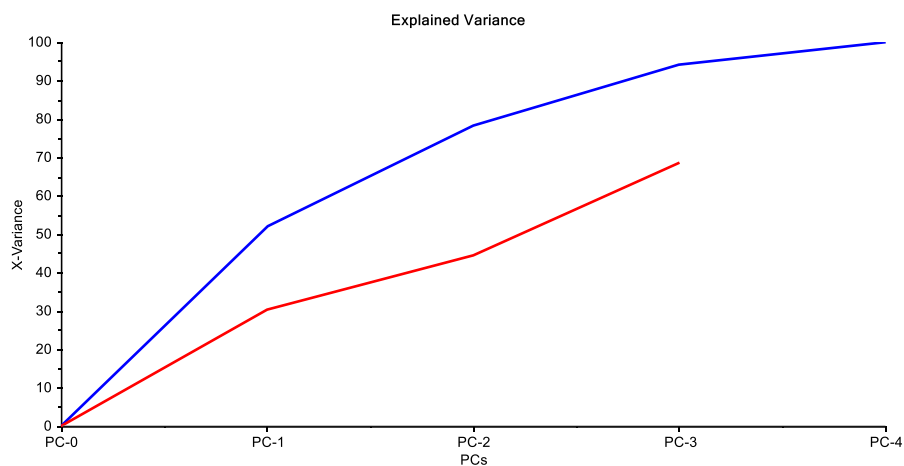
20. Select “NIPALS” circle in the “Algorithm” tab and click “Finish”.



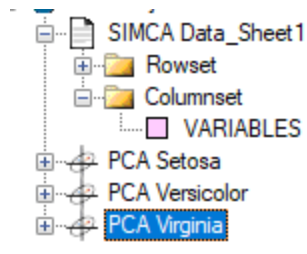
21. Click “Finish” to create PCA model.

22. **Determine** optimal number of components for this PCA model with the explained variance graph.

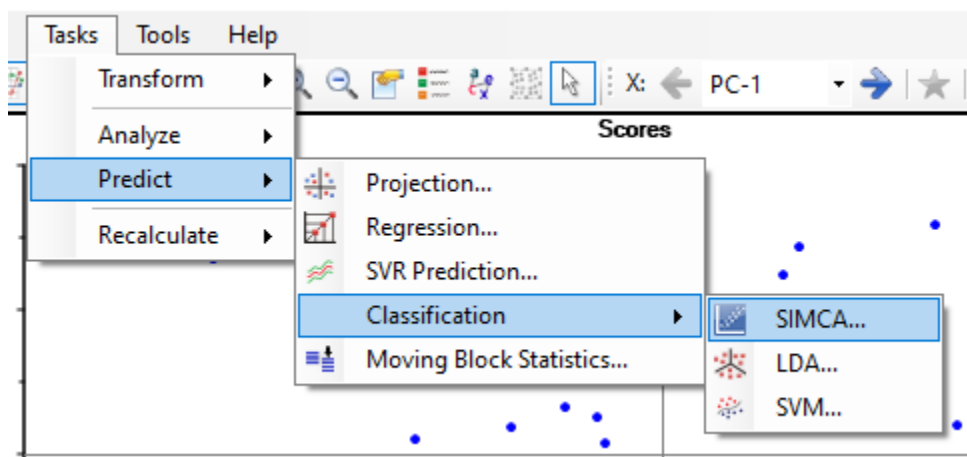
For this example, PC1 would suffice; other components contain more overfitting due to the training (blue) and test (red) lines not being completely parallel.



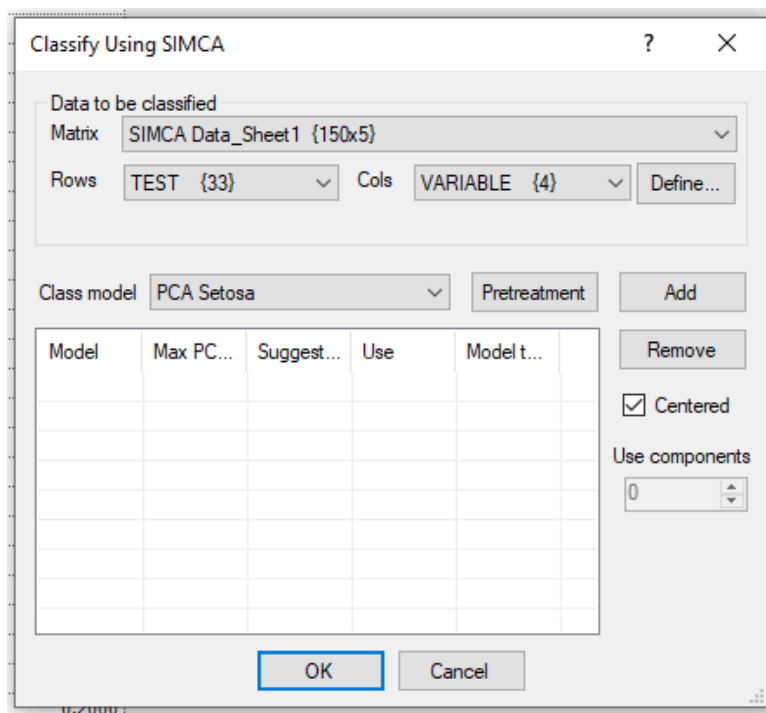
23. **Repeat** steps 15-22 to create PCA models for each class.



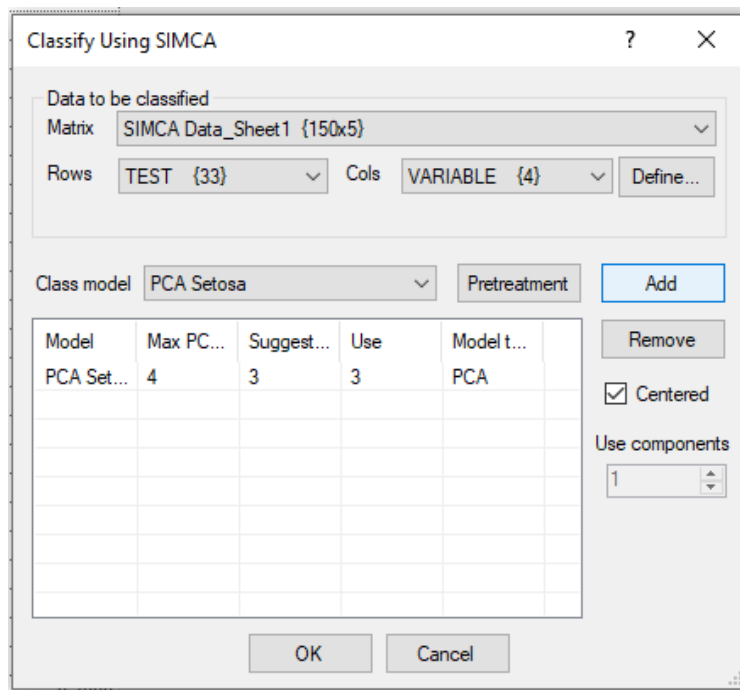
24. **Create** SIMCA model by going to Task → Predict → Classification → SIMCA...



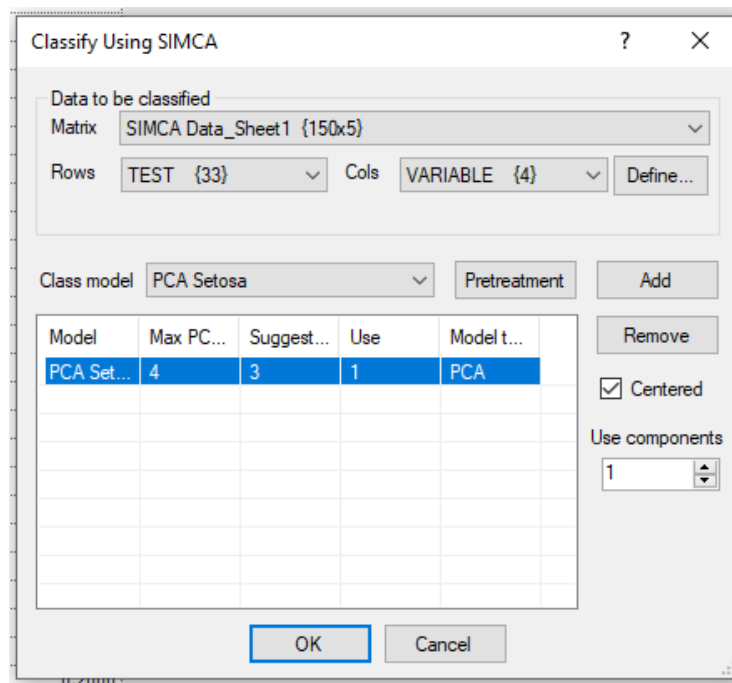
25. **Choose** “VARIABLE” for the Cols set and “TEST” for the Rows set.



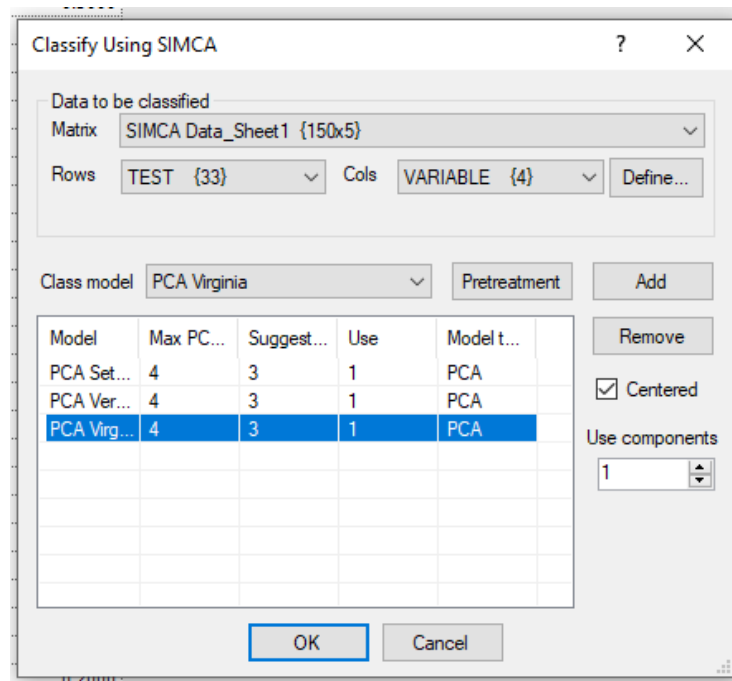
26. **Select** “PCA Setosa” PCA model and click “Add”.



27. **Click** on the “PCA Setosa” model below the “Use” column and place the number of components determined in step 22 for the model using the “Use components” area.



28. **Repeat** steps 26 and 27 for the other PCA models.



29. **Click “OK”** to run SIMCA model.

Sample - Class	PCA Setos	PCA Versi	PCA Virgi
setosa 40	*		
setosa 41	*		
setosa 42	*		
setosa 43	*		
setosa 44	*		
setosa 45	*		
setosa 46	*		
setosa 47	*		
setosa 48	*		
setosa 49	*		
setosa 50	*		
versicolor 40		*	
versicolor 41		*	
versicolor 42		*	*
versicolor 43		*	
versicolor 44		*	
versicolor 45		*	
versicolor 46		*	*
versicolor 47		*	*
versicolor 48		*	*
versicolor 49		*	
versicolor 50		*	*
virginica 40			*
virginica 41			*
virginica 42			*
virginica 43			*
virginica 44			*
virginica 45			*
virginica 46			*
virginica 47			*
virginica 48		*	*
virginica 49			*
virginica 50		*	*

SIMCA shows results through the Classification Table.

Appendix Section

Iris data set used for this example.

Sample	Class	Sepal length	Sepal width	Petal length	Petal width
setosa 1	setosa	5.1	3.5	1.4	0.2
setosa 2	setosa	4.9	3.0	1.4	0.2
setosa 3	setosa	4.7	3.2	1.3	0.2
setosa 4	setosa	4.6	3.1	1.5	0.2
setosa 5	setosa	5.0	3.6	1.4	0.2
setosa 6	setosa	5.4	3.9	1.7	0.4
setosa 7	setosa	4.6	3.4	1.4	0.3
setosa 8	setosa	5.0	3.4	1.5	0.2
setosa 9	setosa	4.4	2.9	1.4	0.2
setosa 10	setosa	4.9	3.1	1.5	0.1
setosa 11	setosa	5.4	3.7	1.5	0.2
setosa 12	setosa	4.8	3.4	1.6	0.2
setosa 13	setosa	4.8	3.0	1.4	0.1
setosa 14	setosa	4.3	3.0	1.1	0.1
setosa 15	setosa	5.8	4.0	1.2	0.2
setosa 16	setosa	5.7	4.4	1.5	0.4
setosa 17	setosa	5.4	3.9	1.3	0.4
setosa 18	setosa	5.1	3.5	1.4	0.3
setosa 19	setosa	5.7	3.8	1.7	0.3
setosa 20	setosa	5.1	3.8	1.5	0.3
setosa 21	setosa	5.4	3.4	1.7	0.2
setosa 22	setosa	5.1	3.7	1.5	0.4
setosa 23	setosa	4.6	3.6	1.0	0.2
setosa 24	setosa	5.1	3.3	1.7	0.5
setosa 25	setosa	4.8	3.4	1.9	0.2
setosa 26	setosa	5.0	3.0	1.6	0.2
setosa 27	setosa	5.0	3.4	1.6	0.4
setosa 28	setosa	5.2	3.5	1.5	0.2
setosa 29	setosa	5.2	3.4	1.4	0.2
setosa 30	setosa	4.7	3.2	1.6	0.2
setosa 31	setosa	4.8	3.1	1.6	0.2
setosa 32	setosa	5.4	3.4	1.5	0.4
setosa 33	setosa	5.2	4.1	1.5	0.1
setosa 34	setosa	5.5	4.2	1.4	0.2
setosa 35	setosa	4.9	3.1	1.5	0.2
setosa 36	setosa	5.0	3.2	1.2	0.2
setosa 37	setosa	5.5	3.5	1.3	0.2
setosa 38	setosa	4.9	3.6	1.4	0.1
setosa 39	setosa	4.4	3.0	1.3	0.2
setosa 40	setosa	5.1	3.4	1.5	0.2
setosa 41	setosa	5.0	3.5	1.3	0.3
setosa 42	setosa	4.5	2.3	1.3	0.3
setosa 43	setosa	4.4	3.2	1.3	0.2
setosa 44	setosa	5.0	3.5	1.6	0.6
setosa 45	setosa	5.1	3.8	1.9	0.4
setosa 46	setosa	4.8	3.0	1.4	0.3
setosa 47	setosa	5.1	3.8	1.6	0.2
setosa 48	setosa	4.6	3.2	1.4	0.2
setosa 49	setosa	5.3	3.7	1.5	0.2
setosa 50	setosa	5.0	3.3	1.4	0.2
versicolor 1	versicolor	7.0	3.2	4.7	1.4
versicolor 2	versicolor	6.4	3.2	4.5	1.5
versicolor 3	versicolor	6.9	3.1	4.9	1.5
versicolor 4	versicolor	5.5	2.3	4.0	1.3
versicolor 5	versicolor	6.5	2.8	4.6	1.5
versicolor 6	versicolor	5.7	2.8	4.5	1.3
versicolor 7	versicolor	6.3	3.3	4.7	1.6
versicolor 8	versicolor	4.9	2.4	3.3	1.0
versicolor 9	versicolor	6.6	2.9	4.6	1.3
versicolor 10	versicolor	5.2	2.7	3.9	1.4

versicolor 11	versicolor	5.0	2.0	3.5	1.0
versicolor 12	versicolor	5.9	3.0	4.2	1.5
versicolor 13	versicolor	6.0	2.2	4.0	1.0
versicolor 14	versicolor	6.1	2.9	4.7	1.4
versicolor 15	versicolor	5.6	2.9	3.6	1.3
versicolor 16	versicolor	6.7	3.1	4.4	1.4
versicolor 17	versicolor	5.6	3.0	4.5	1.5
versicolor 18	versicolor	5.8	2.7	4.1	1.0
versicolor 19	versicolor	6.2	2.2	4.5	1.5
versicolor 20	versicolor	5.6	2.5	3.9	1.1
versicolor 21	versicolor	5.9	3.2	4.8	1.8
versicolor 22	versicolor	6.1	2.8	4.0	1.3
versicolor 23	versicolor	6.3	2.5	4.9	1.5
versicolor 24	versicolor	6.1	2.8	4.7	1.2
versicolor 25	versicolor	6.4	2.9	4.3	1.3
versicolor 26	versicolor	6.6	3.0	4.4	1.4
versicolor 27	versicolor	6.8	2.8	4.8	1.4
versicolor 28	versicolor	6.7	3.0	5.0	1.7
versicolor 29	versicolor	6.0	2.9	4.5	1.5
versicolor 30	versicolor	5.7	2.6	3.5	1.0
versicolor 31	versicolor	5.5	2.4	3.8	1.1
versicolor 32	versicolor	5.5	2.4	3.7	1.0
versicolor 33	versicolor	5.8	2.7	3.9	1.2
versicolor 34	versicolor	6.0	2.7	5.1	1.6
versicolor 35	versicolor	5.4	3.0	4.5	1.5
versicolor 36	versicolor	6.0	3.4	4.5	1.6
versicolor 37	versicolor	6.7	3.1	4.7	1.5
versicolor 38	versicolor	6.3	2.3	4.4	1.3
versicolor 39	versicolor	5.6	3.0	4.1	1.3
versicolor 40	versicolor	5.5	2.5	4.0	1.3
versicolor 41	versicolor	5.5	2.6	4.4	1.2
versicolor 42	versicolor	6.1	3.0	4.6	1.4
versicolor 43	versicolor	5.8	2.6	4.0	1.2
versicolor 44	versicolor	5.0	2.3	3.3	1.0
versicolor 45	versicolor	5.6	2.7	4.2	1.3
versicolor 46	versicolor	5.7	3.0	4.2	1.2
versicolor 47	versicolor	5.7	2.9	4.2	1.3
versicolor 48	versicolor	6.2	2.9	4.3	1.3
versicolor 49	versicolor	5.1	2.5	3.0	1.1
versicolor 50	versicolor	5.7	2.8	4.1	1.3
virginica 1	virginica	6.3	3.3	6.0	2.5
virginica 2	virginica	5.8	2.7	5.1	1.9
virginica 3	virginica	7.1	3.0	5.9	2.1
virginica 4	virginica	6.3	2.9	5.6	1.8
virginica 5	virginica	6.5	3.0	5.8	2.2
virginica 6	virginica	7.6	3.0	6.6	2.1
virginica 7	virginica	4.9	2.5	4.5	1.7
virginica 8	virginica	7.3	2.9	6.3	1.8
virginica 9	virginica	6.7	2.5	5.8	1.8
virginica 10	virginica	7.2	3.6	6.1	2.5
virginica 11	virginica	6.5	3.2	5.1	2.0
virginica 12	virginica	6.4	2.7	5.3	1.9
virginica 13	virginica	6.8	3.0	5.5	2.1
virginica 14	virginica	5.7	2.5	5.0	2.0
virginica 15	virginica	5.8	2.8	5.1	2.4
virginica 16	virginica	6.4	3.2	5.3	2.3
virginica 17	virginica	6.5	3.0	5.5	1.8
virginica 18	virginica	7.7	3.8	6.7	2.2
virginica 19	virginica	7.7	2.6	6.9	2.3
virginica 20	virginica	6.0	2.2	5.0	1.5
virginica 21	virginica	6.9	3.2	5.7	2.3
virginica 22	virginica	5.6	2.8	4.9	2.0
virginica 23	virginica	7.7	2.8	6.7	2.0
virginica 24	virginica	6.3	2.7	4.9	1.8
virginica 25	virginica	6.7	3.3	5.7	2.1
virginica 26	virginica	7.2	3.2	6.0	1.8

virginica 27	virginica	6.2	2.8	4.8	1.8
virginica 28	virginica	6.1	3.0	4.9	1.8
virginica 29	virginica	6.4	2.8	5.6	2.1
virginica 30	virginica	7.2	3.0	5.8	1.6
virginica 31	virginica	7.4	2.8	6.1	1.9
virginica 32	virginica	7.9	3.8	6.4	2.0
virginica 33	virginica	6.4	2.8	5.6	2.2
virginica 34	virginica	6.3	2.8	5.1	1.5
virginica 35	virginica	6.1	2.6	5.6	1.4
virginica 36	virginica	7.7	3.0	6.1	2.3
virginica 37	virginica	6.3	3.4	5.6	2.4
virginica 38	virginica	6.4	3.1	5.5	1.8
virginica 39	virginica	6.0	3.0	4.8	1.8
virginica 40	virginica	6.9	3.1	5.4	2.1
virginica 41	virginica	6.7	3.1	5.6	2.4
virginica 42	virginica	6.9	3.1	5.1	2.3
virginica 43	virginica	5.8	2.7	5.1	1.9
virginica 44	virginica	6.8	3.2	5.9	2.3
virginica 45	virginica	6.7	3.3	5.7	2.5
virginica 46	virginica	6.7	3.0	5.2	2.3
virginica 47	virginica	6.3	2.5	5.0	1.9
virginica 48	virginica	6.5	3.0	5.2	2.0
virginica 49	virginica	6.2	3.4	5.4	2.3
virginica 50	virginica	5.9	3.0	5.1	1.8

Advisor Signature

Co-Advisor Signature
