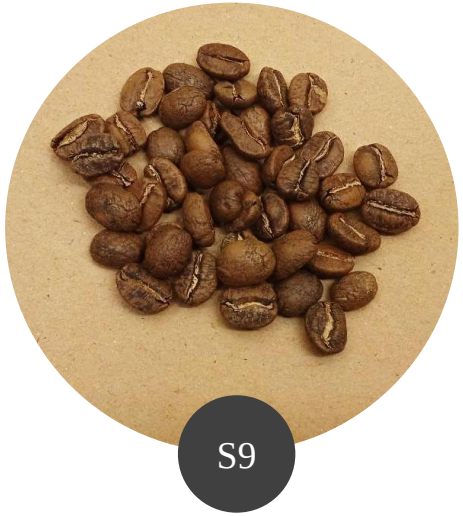
The Use of PLS-DA Method and NIR Spectroscopy for Classification of Robusta and Arabica Coffee Beans

Introduction

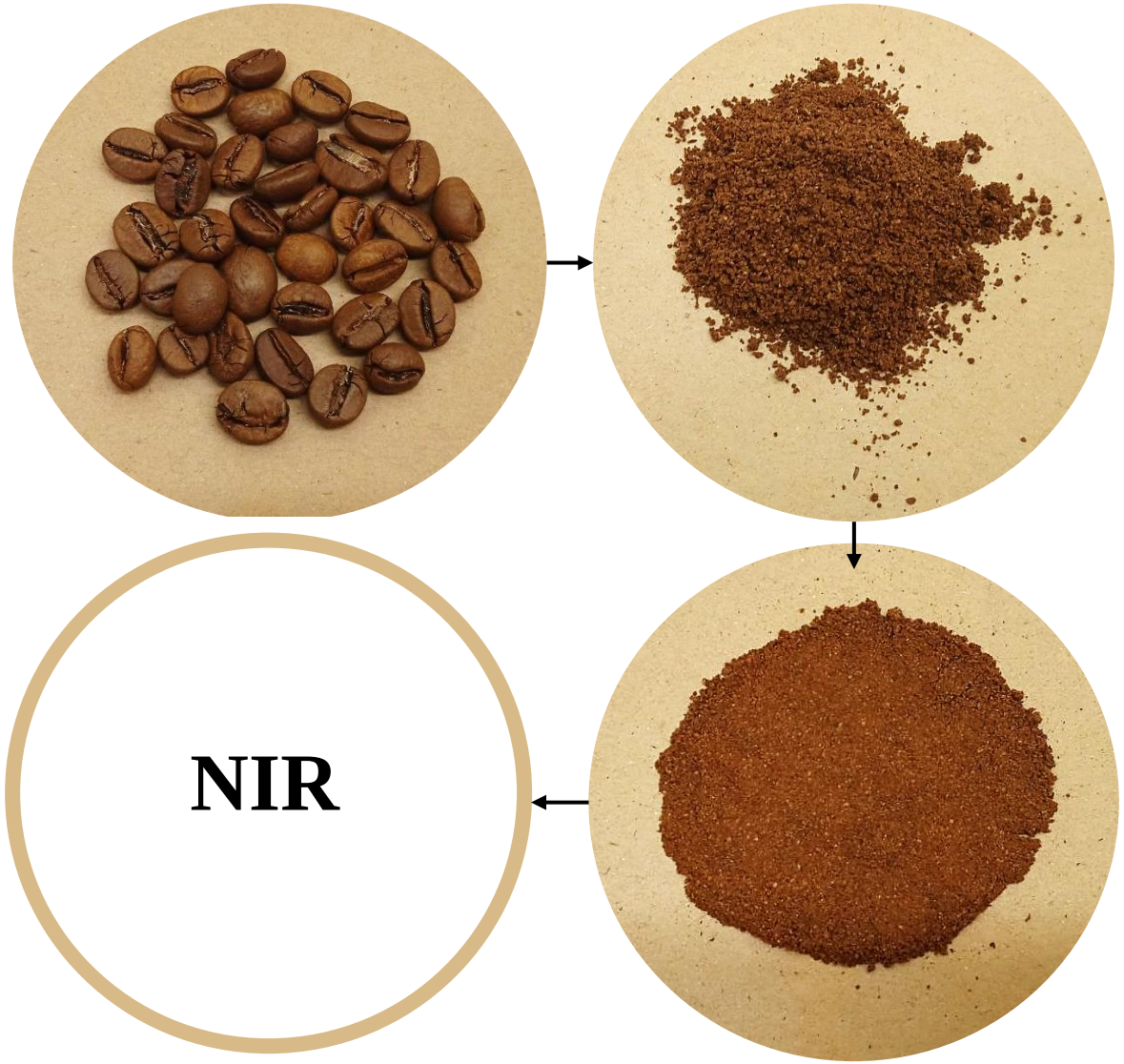
- Coffee is one of the most popular beverages in the world and especially in Iran. Generally, the seeds of this plant are not cultivated in Iran and most of the country's consumption is provided through roasting and processing of imported coffee beans. The main goal of this study is development of suitable and quick procedures based on benchtop and handheld NIR spectroscopy by using *Partial Least Squares-Discriminant Analysis (PLS-DA)* method for analysis of spectral data and classification of coffee Beans available in Iran. One of the applications of such studies could be in coffee authentication with low-cost analytical methods. In these applications, it is considerable to use a handheld NIR spectrometer.

- Despite containing less caffeine than Robusta, Arabica beans are often considered superior in taste. Arabica tends to have a smoother, sweeter taste, with flavour notes of chocolate and sugar. They often also have hints of fruits or berries. Robusta, on the other hand, has a stronger, harsher and more bitter taste, with grainy or rubbery overtones.

Samples

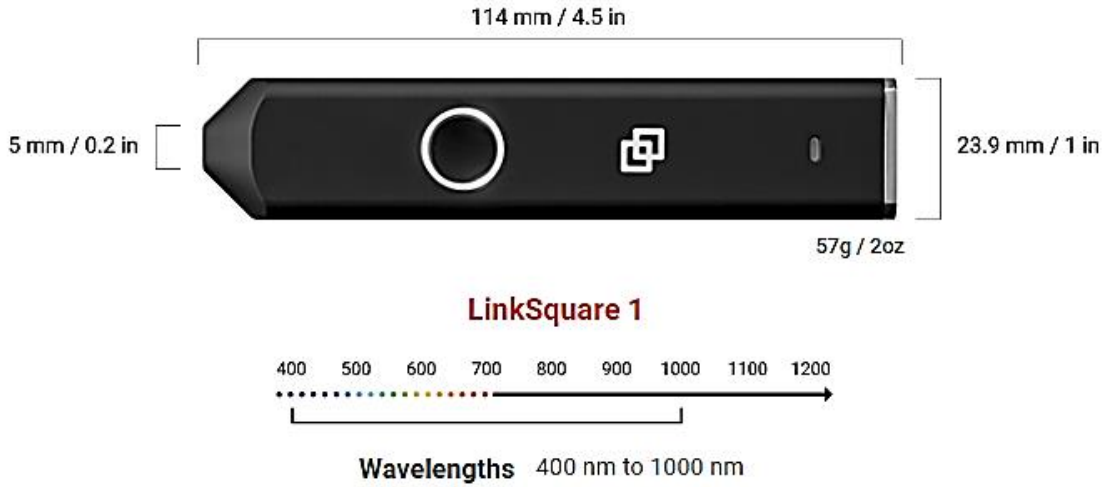


Sample Preparation

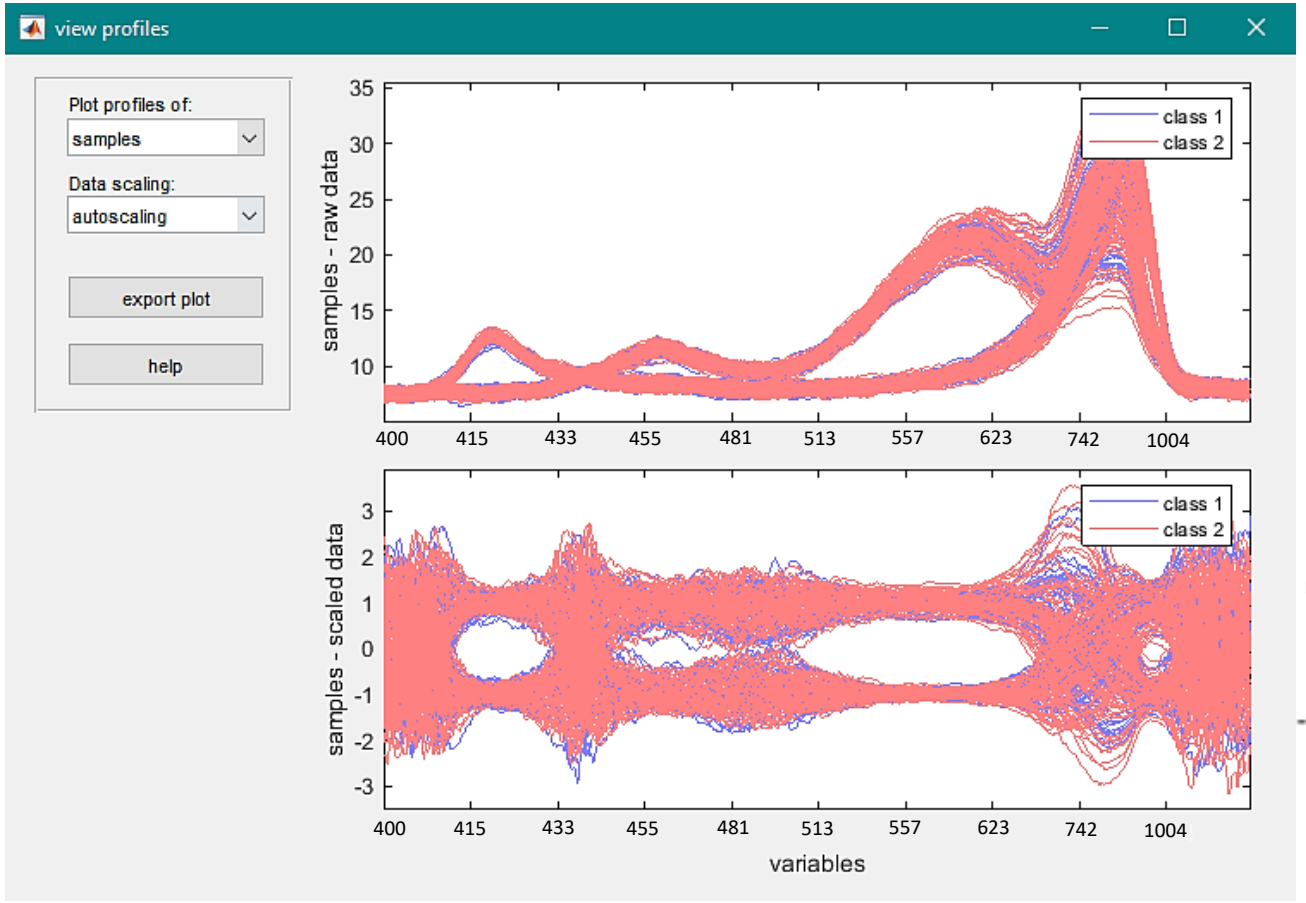


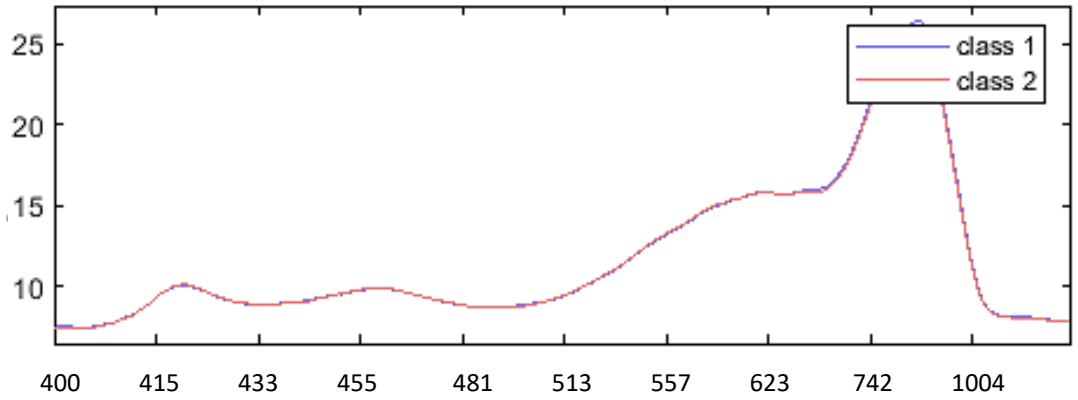
Sample	Type	Region
S1	Robusta	Uganda
S2	Robusta	Vietnam, UM
S3	Robusta	Indonesia, Java
S4	Robusta	India, Peaberry
S5	Robusta	India, Cherry
S6	Arabica	Etiyopya
S7	Arabica	Brasil, Rio
S8	Arabica	Honduras
S9	Arabica	Colombia, Tabilla
S10	Arabica	Colombia, Supremo

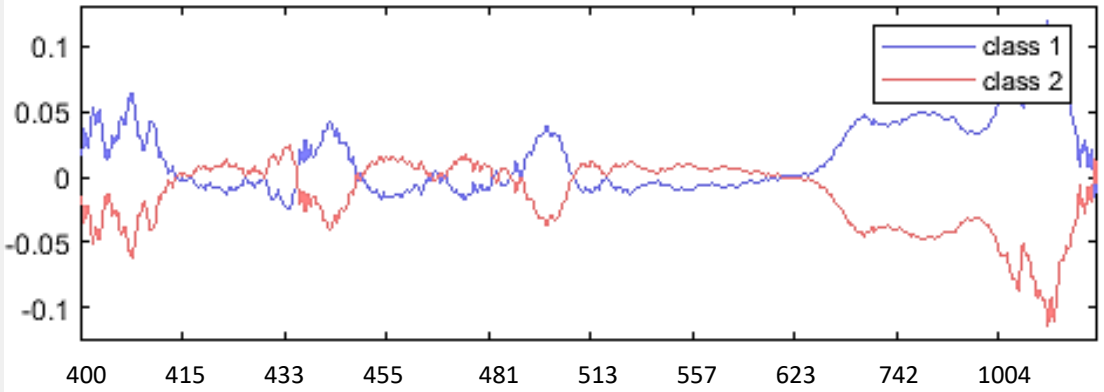
NIR Analysis



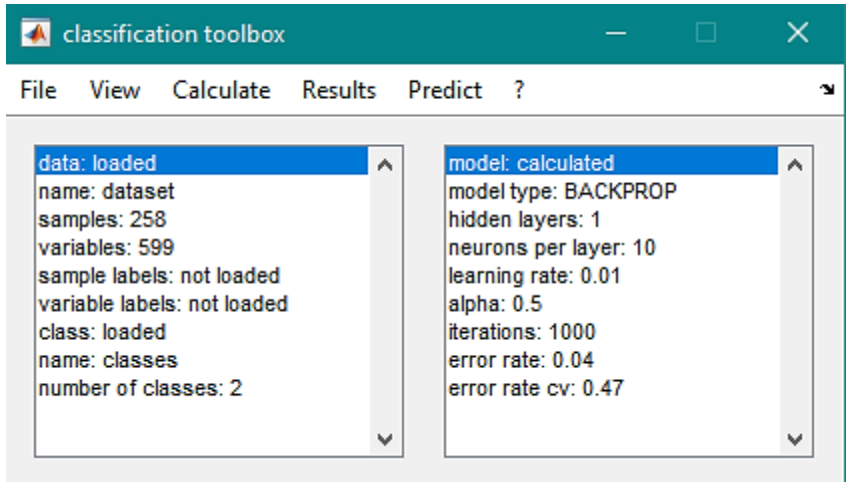
Pre-Processing

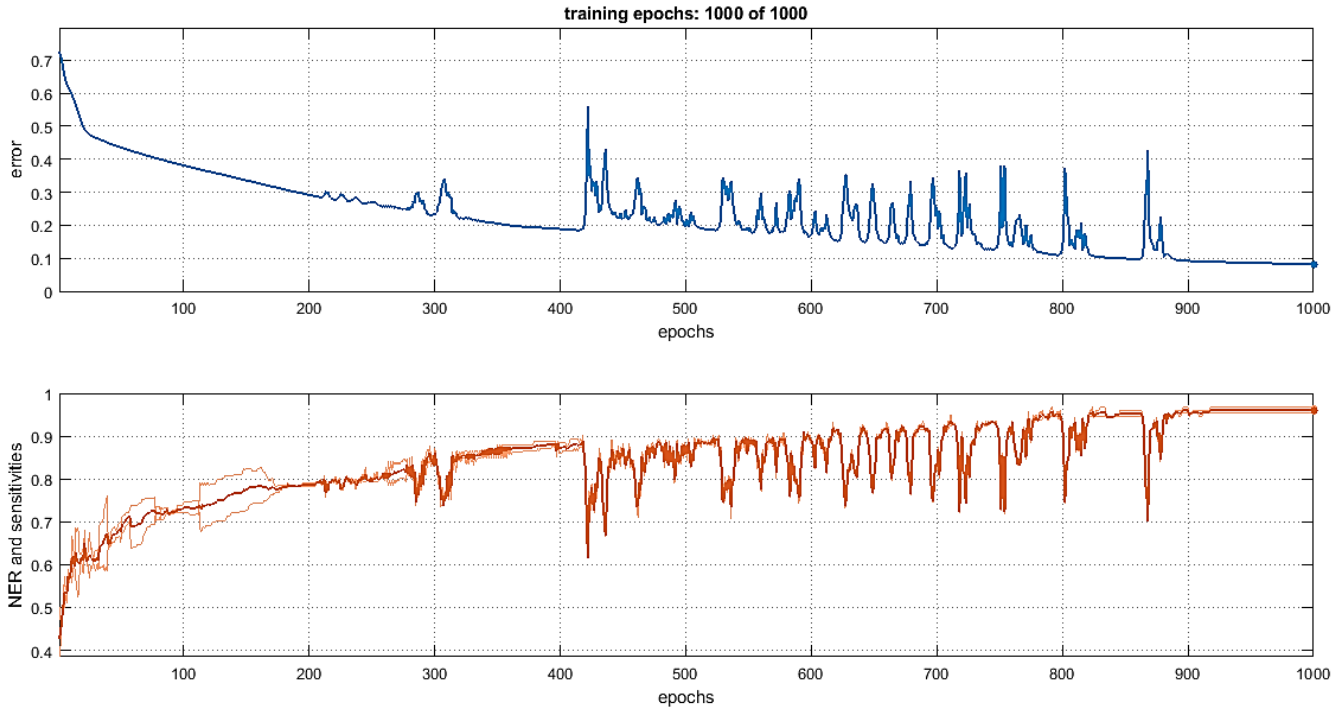




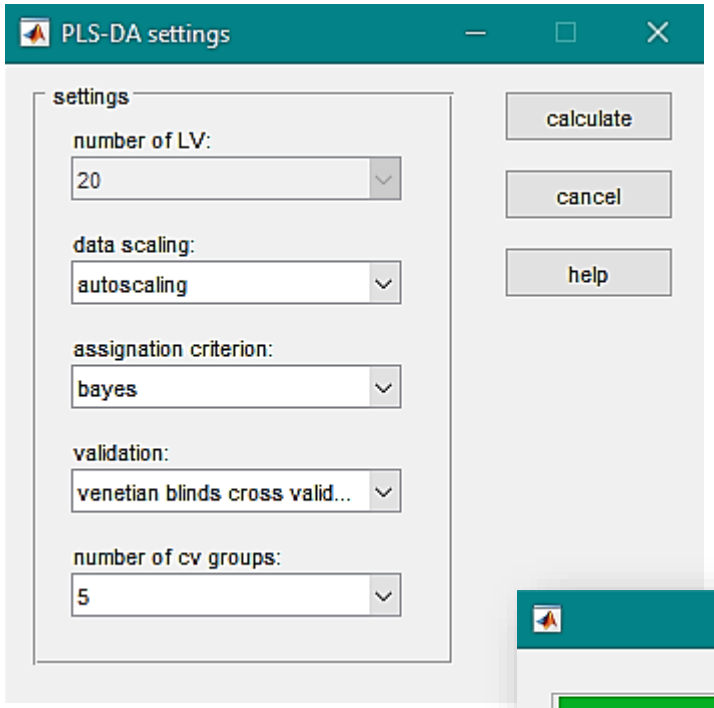


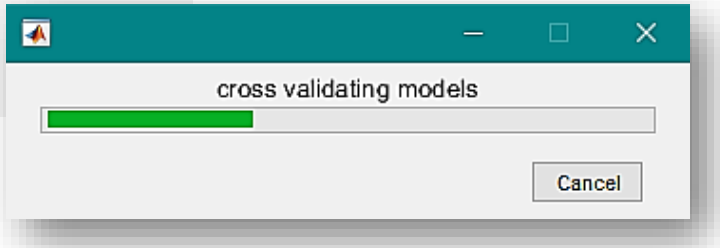
Back Propagation

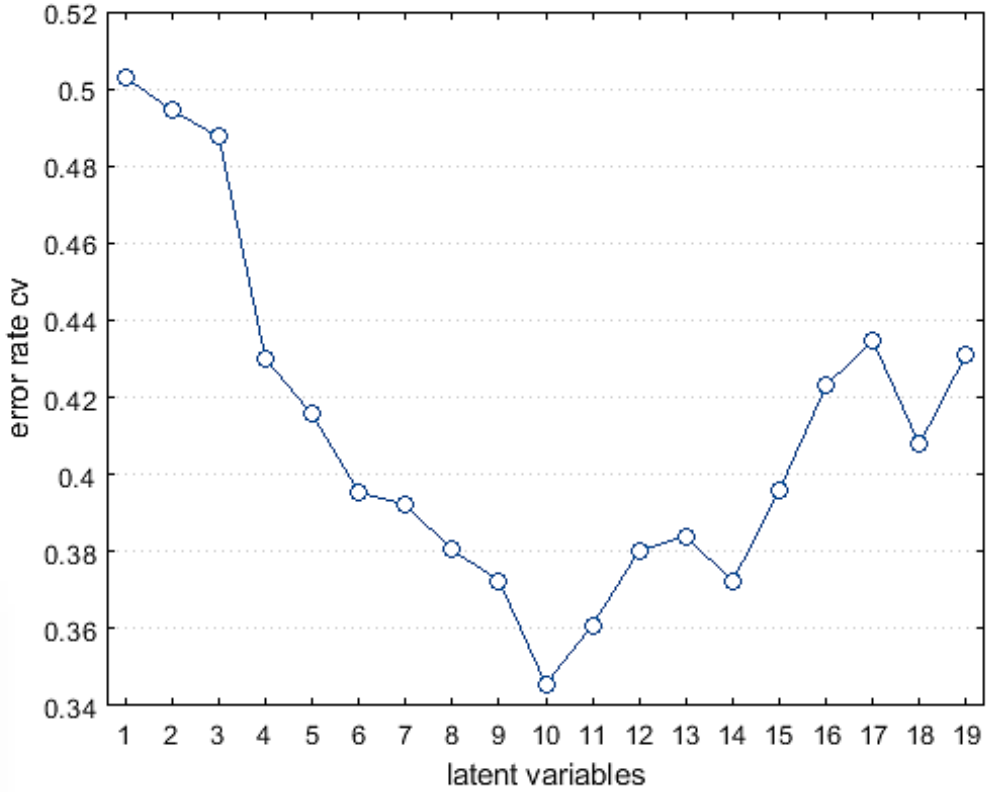




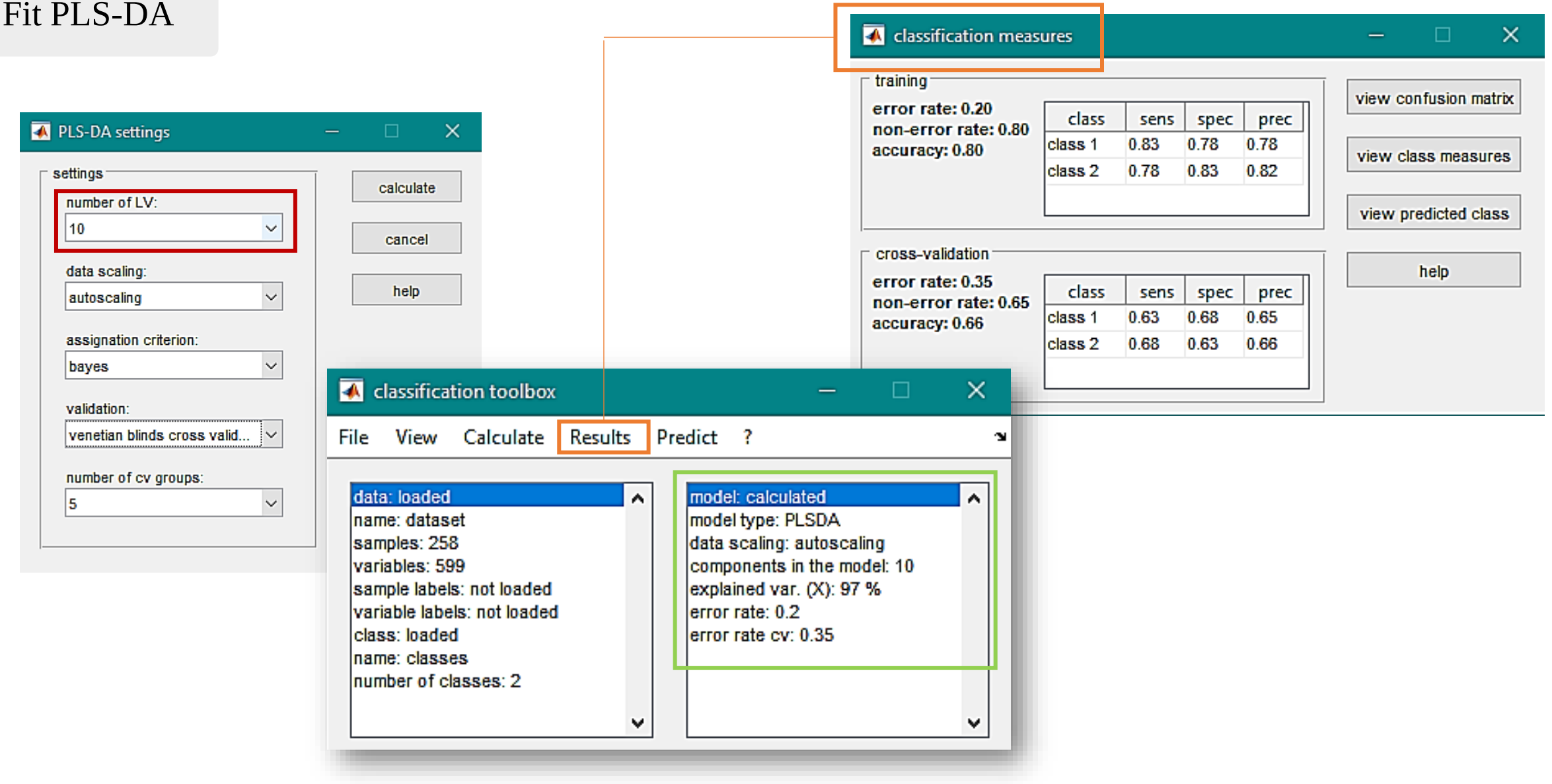
Optimal Component for PLS-DA



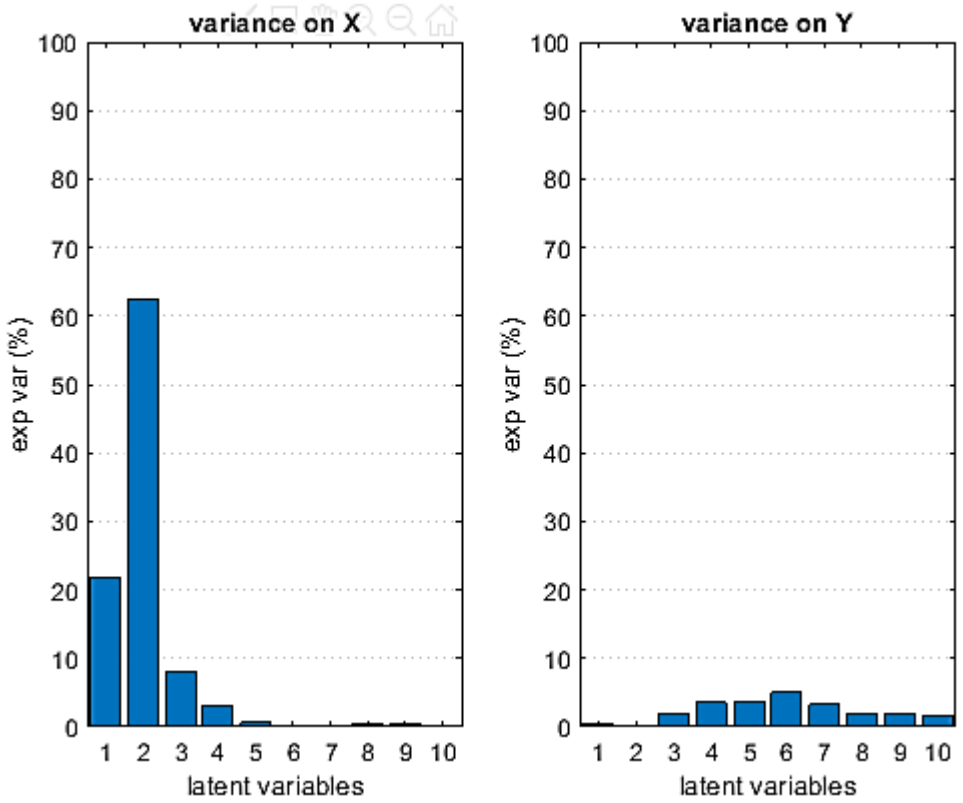


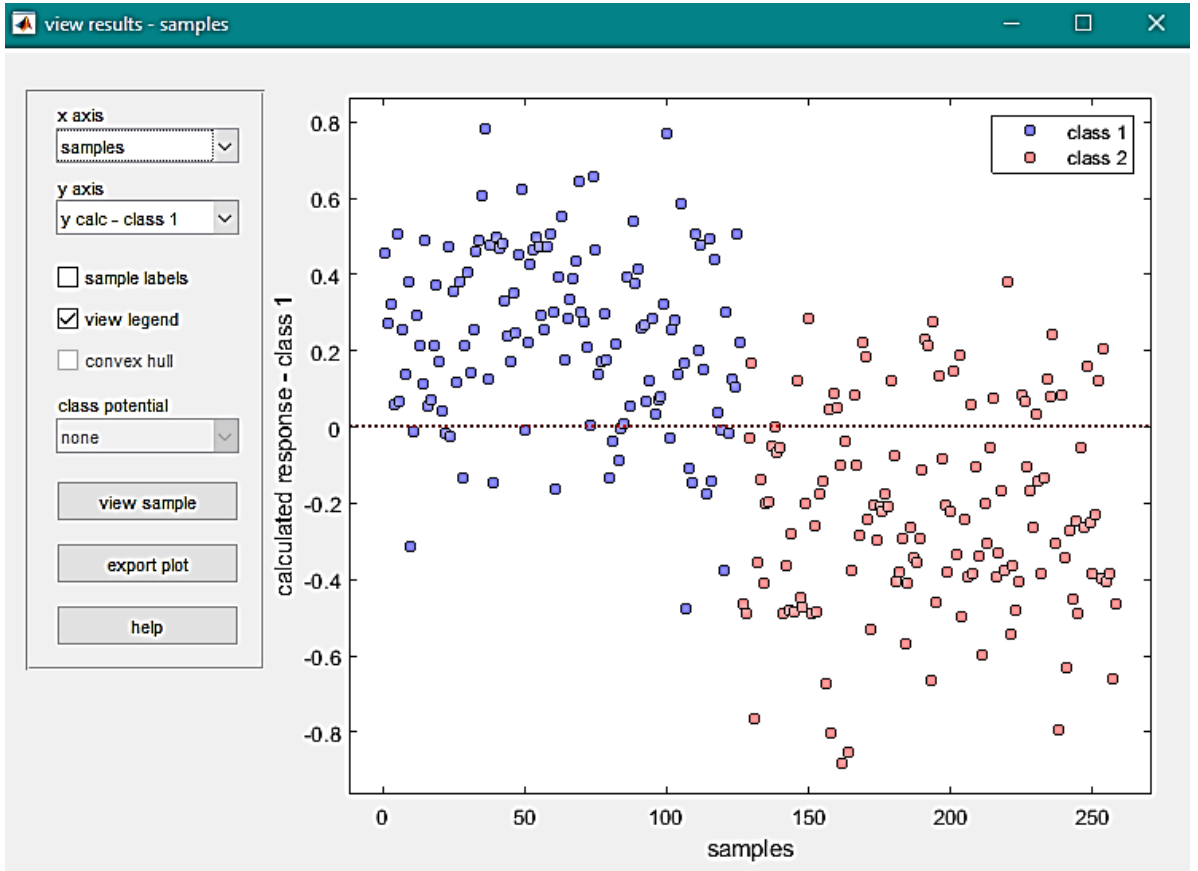


Fit PLS-DA

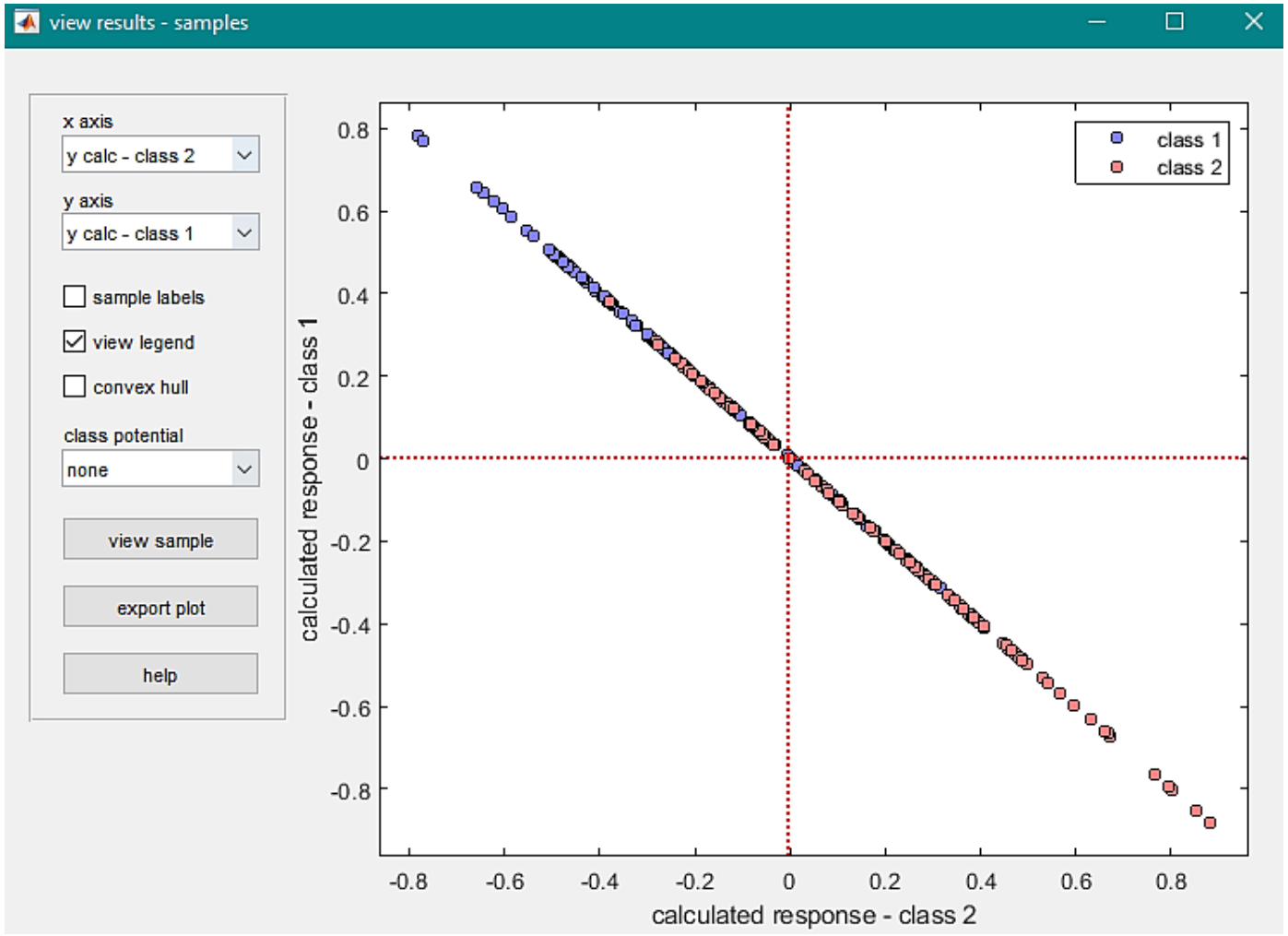

PLS-DA settings
settings
number of LV:
10
data scaling:
autoscaling
assignation criterion:
bayes
validation:
venetian blinds cross valid...
number of cv groups:
5
calculate
cancel
help
classification measures
training
error rate: 0.20
non-error rate: 0.80
accuracy: 0.80
class sens spec prec
class 1 0.83 0.78 0.78
class 2 0.78 0.83 0.82
cross-validation
error rate: 0.35
non-error rate: 0.65
accuracy: 0.66
class sens spec prec
class 1 0.63 0.68 0.65
class 2 0.68 0.63 0.66
view confusion matrix
view class measures
view predicted class
help
classification toolbox
File View Calculate Results Predict ?
data: loaded
name: dataset
samples: 258
variables: 599
sample labels: not loaded
variable labels: not loaded
class: loaded
name: classes
number of classes: 2
model: calculated
model type: PLSDA
data scaling: autoscaling
components in the model: 10
explained var. (X): 97 %
error rate: 0.2
error rate cv: 0.35

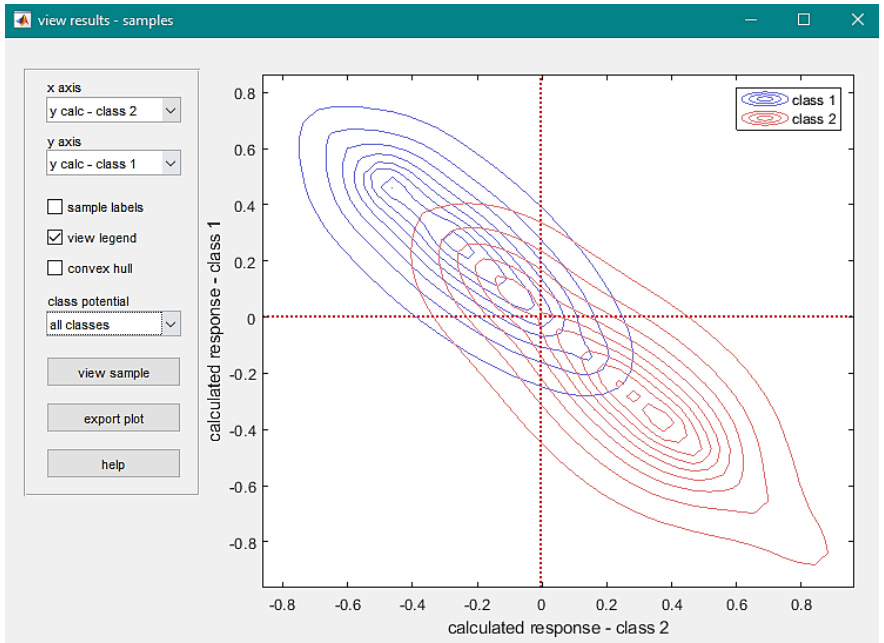
PLS-DA Results



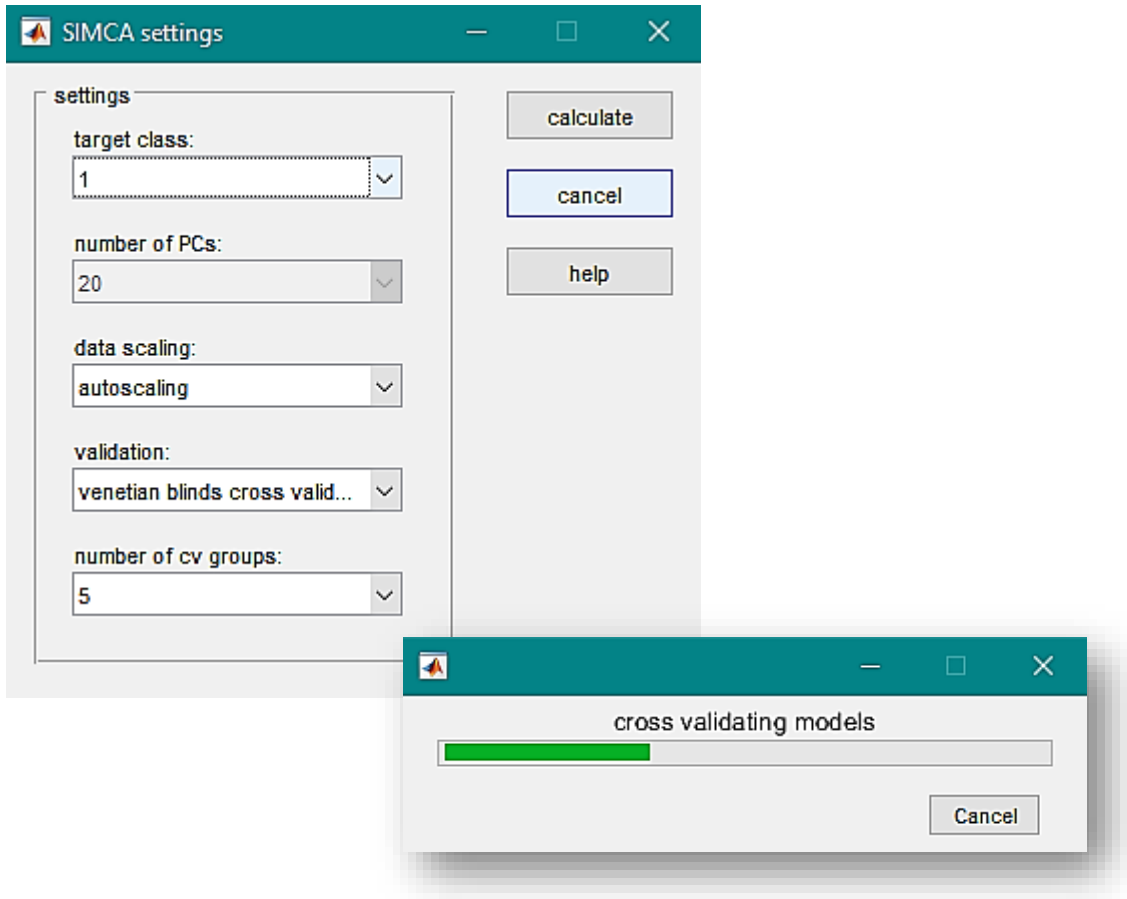


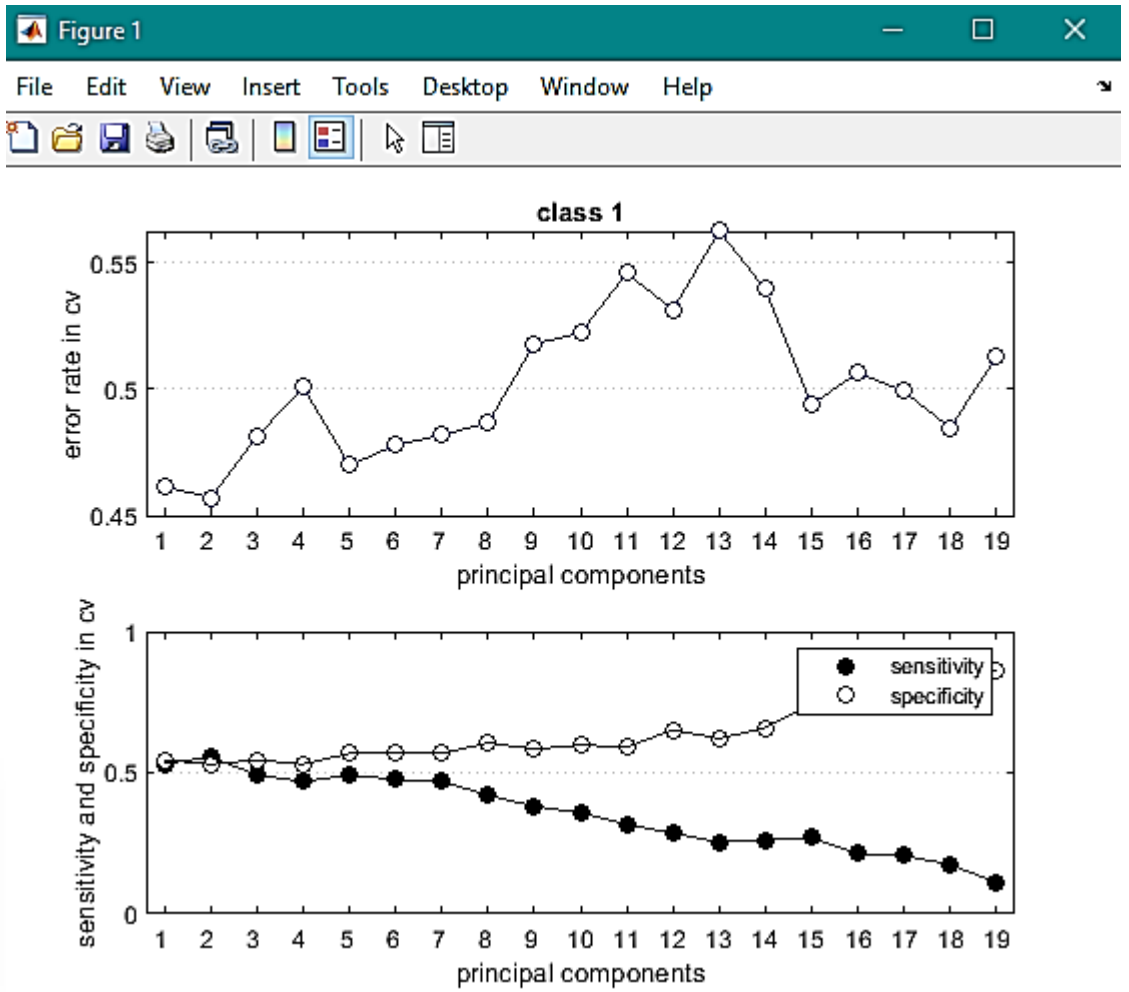
PLS-DA Results



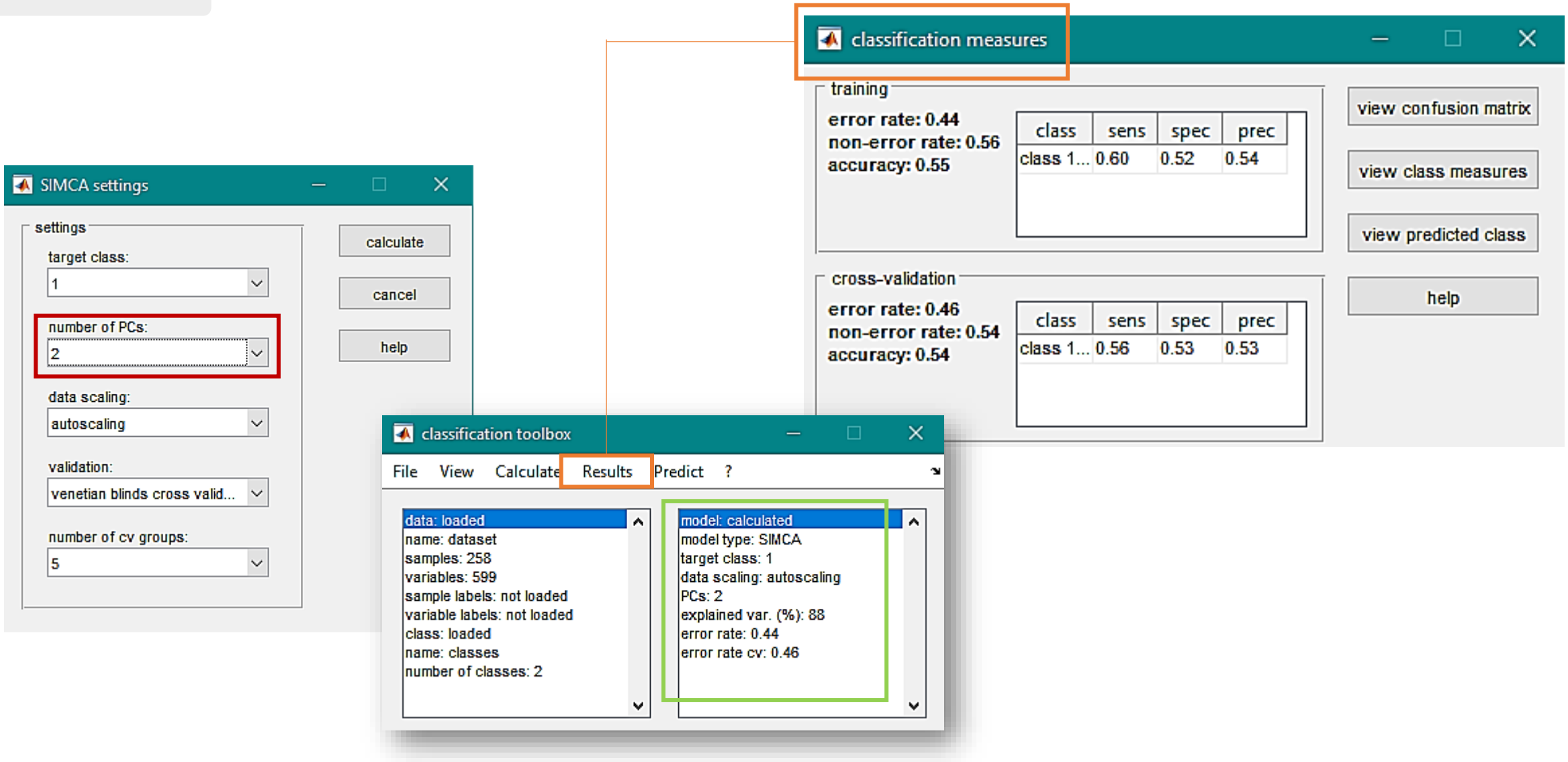


Optimal Component for SIMCA

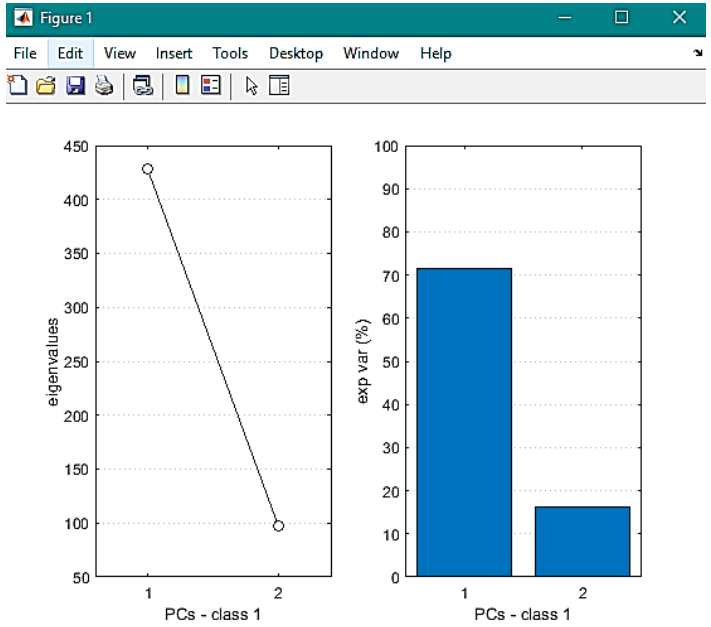


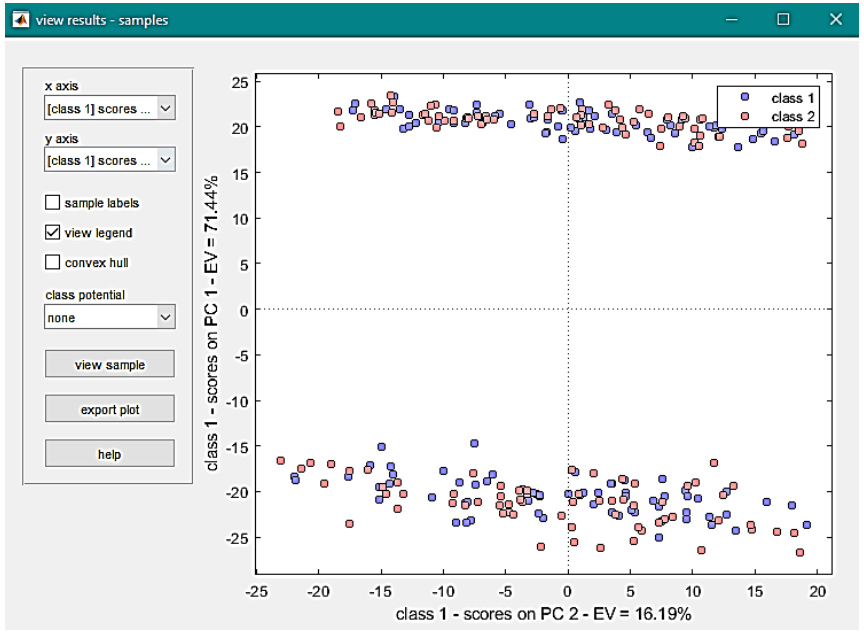


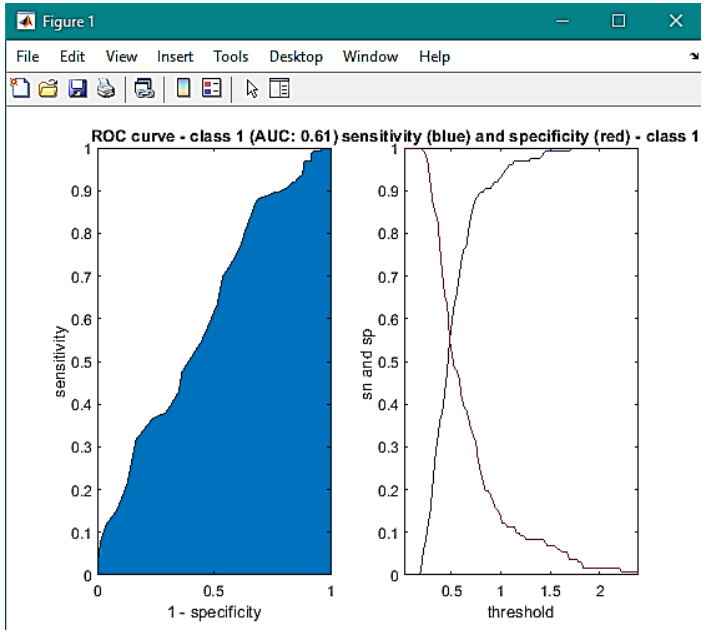
Fit SIMCA


classification measures
training
error rate: 0.44
non-error rate: 0.56
accuracy: 0.55
class | sens | spec | prec
class 1... | 0.60 | 0.52 | 0.54
cross-validation
error rate: 0.46
non-error rate: 0.54
accuracy: 0.54
class | sens | spec | prec
class 1... | 0.56 | 0.53 | 0.53
view confusion matrix
view class measures
view predicted class
help
SIMCA settings
settings
target class:
1
number of PCs:
2
data scaling:
autoscaling
validation:
venetian blinds cross valid...
number of cv groups:
5
calculate
cancel
help
classification toolbox
File
View
Calculate
Results
Predict
?
data: loaded
name: dataset
samples: 258
variables: 599
sample labels: not loaded
variable labels: not loaded
class: loaded
name: classes
number of classes: 2
model: calculated
model type: SIMCA
target class: 1
data scaling: autoscaling
PCs: 2
explained var. (%): 88
error rate: 0.44
error rate cv: 0.46

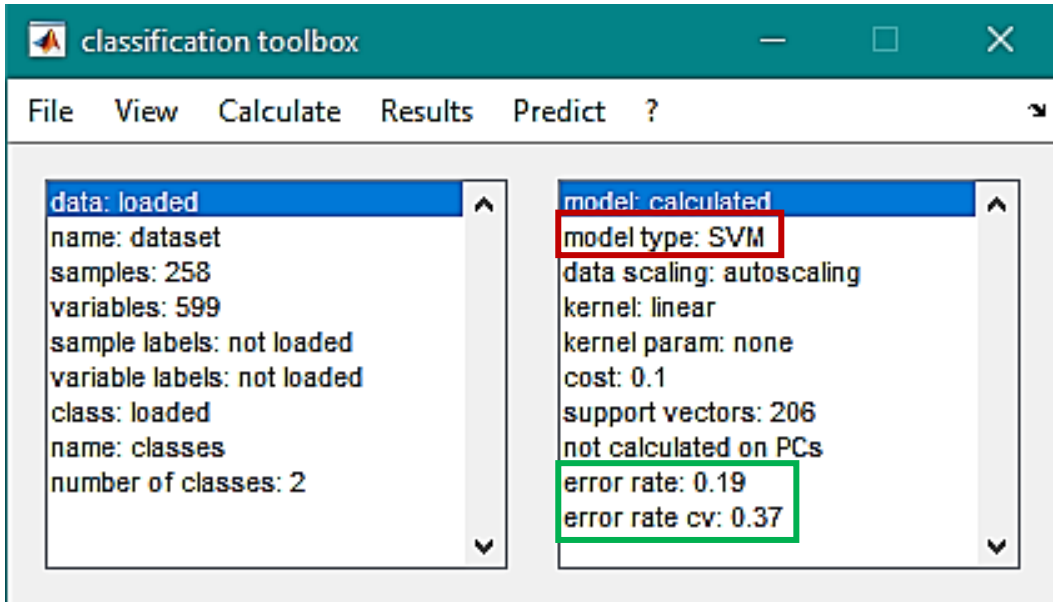
SIMCA Results

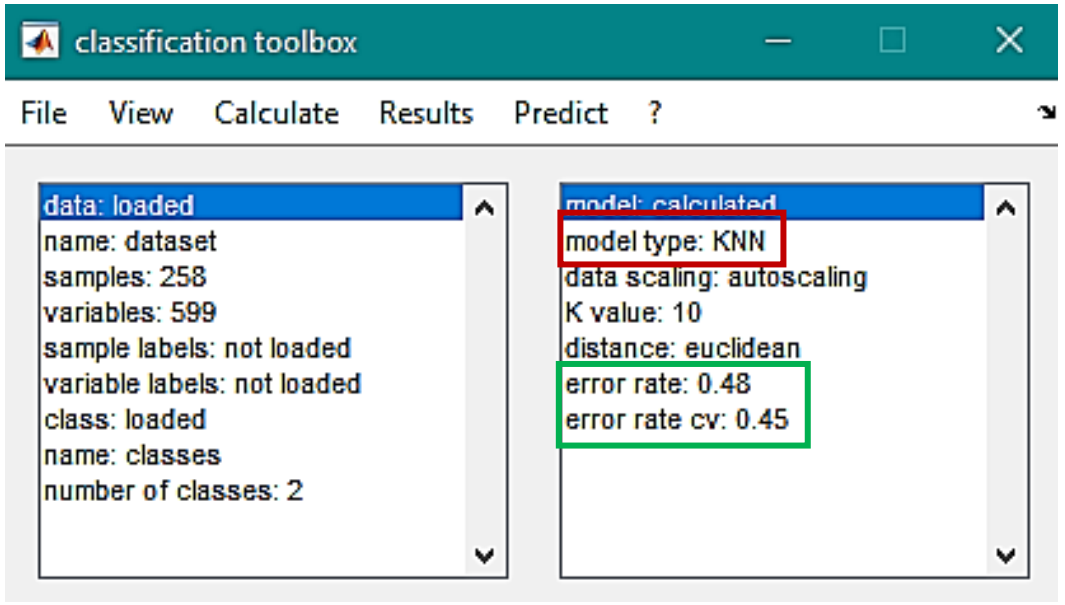






Other Methods





Thank You
For Your Attention!