Moscow Aviation Institute



Exploring data analysis. Preprocessing of data. Forming of feature space and a decrease in its dimension by the method of principal components. Creation of the decisive rule by means of different support (reference) vector machines

Fundamentals of Machine Learning

Associate Professor: Pavel Kudryavtsev

М30-420-БКИ

Student: Ehab Abdelghany

METHODICAL INSTRUCTIONS FOR THE EXECUTION OF LABORATORY WORKS IN THE COURSE "MACHINE LEARNING".

LABORATORY WORK NO. 1.

"Exploring data analysis. Preprocessing of data. Forming of feature space and a decrease in its dimension by the method of principal component. Creation of the decisive rule by means of different support (reference) vector machines".

Work execution order.

1. Receive from the teacher a task. Study the material stated in methodical instructions for laboratory work. Browse an example of the execution of laboratory work.
2. Cause the problem of classification (by means of the start of the Lab1main program) corresponding to a task for the execution of laboratory work. Analyze the structure of character space (covariation matrices of different classes of objects, histograms of signs). Conclude the possibility of application of Bayes classifiers. Conclude an informative feature set.
2. Call the procedure main a component for the purpose of decreasing in character space. Conclude the possibility of considering the two-dimensional space of the combined signs to solve a problem of classification. Carry out the analysis and select two initial signs, the most corresponding to the first two main components.
3. Call the procedure for the start of the main machines of reference vectors. Train machines at the set sample using the first two main components. Analyze the results of the operation of different machines of reference vectors by assessing of errors of reclassification and with use of a control sample.
4. Select the best machine of support vectors for classification of the given training sample of the generalized signs received based on a method main component and a control sample.
5. Call the procedure of the start of the main machines of reference vectors. Train machines at the set sample with the use of the first two main component. Analyze the results of the operation of different machines of reference vectors by assessment of errors of reclassification, and with the use of a control sample. Select the best model of classification.
6. Call the procedure for the start of the main machines of reference vectors. Train machineonat the set sample with the use of the first two informative signs of initial character space. Analyze results of operation of different machines of reference vectors by assessment of errors of reclassification and with the use of a control sample. Select the best model of classification.
7. Compare results of the operation of machines of the reference vectors constructed on space main a component and space of informative signs.
8. Fill out the report.

Enter your var number jvar=9
Cov matrices 1-t and 2-d classes

RdatNr1 =

0.9027	0.0816	-0.2783	-0.4370
0.0816	2.4064	0.0703	-0.4419
-0.2783	0.0703	1.3191	0.0981
-0.4370	-0.4419	0.0981	2.8277

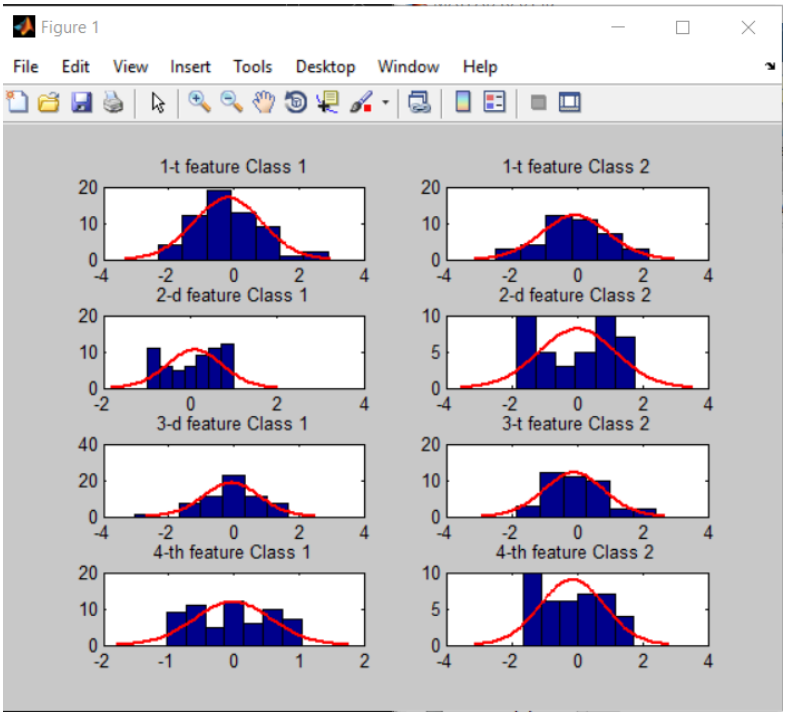
RdatNr2 =

0.6308	-0.0145	0.1318	0.0458
-0.0145	0.4683	-0.1158	0.0548
0.1318	-0.1158	0.7584	-0.0350
0.0458	0.0548	-0.0350	0.6727

General assessment of sample covariation matrix

RdatNrN =

0.9295	0.0061	-0.0589	-0.0669
0.0061	0.6964	-0.0614	0.0085
-0.0589	-0.0614	0.6871	-0.0012
-0.0669	0.0085	-0.0012	0.5203



Analyse results, and make conclusions. Save results and press any key to continue
Eigenvalues of general covariation matrix

ans =

0.5088
0.6250
0.7453
0.9541

Coeficcients of transp PCA component

ans =

0.9600	0.0722	-0.2279	-0.1461
-0.1895	0.7615	-0.6135	0.0882
0.1252	0.6432	0.7545	-0.0364
0.1640	-0.0338	0.0490	0.9847

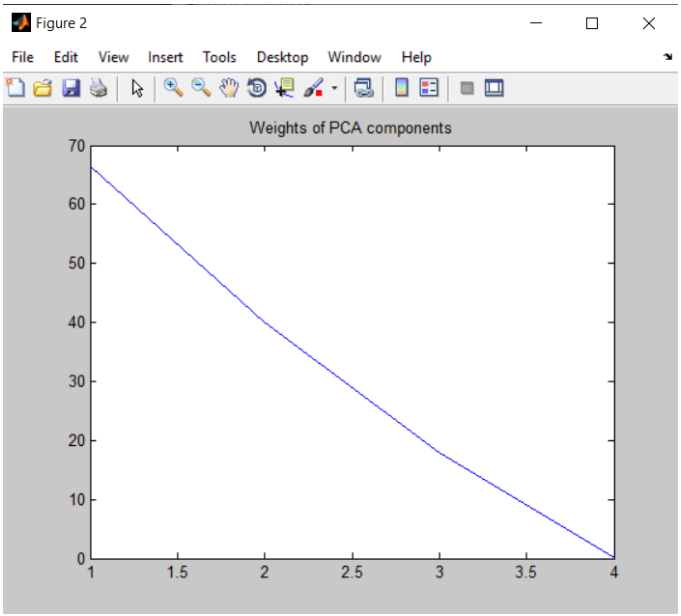
Explanation of choosing PCA weights

Explan =

33.6761
26.3070
22.0602
17.9567

Analysis using main PCA components

NumPCA =2



ake analysis matrices Coef,Explan and choose inform coord
Enter the first information coord number ninf1=3
Enter the second information coord number ninf2=1
Make analysis using 2 main PCA components
Running time (CPU secs) = 0
Number of iterations = 1
Training accuracy = 0.59

iter =

1

wlin =

0.0325
-0.0298

gammalin =

0.0967

Lin SVM regr err in linear space
Lin SVM contr err in lin space
Running time (CPU secs) = 0
Number of iterations = 1
Training accuracy = 0.7

wKM =

-0.0872
0.1458

gammaKM =

0.1152

Lin SVM regrated error in quadratic space
Lin SVM test error in quadratic space
SVM regrated error with Gauss core
SVM test error with Gauss core
SVM regrated error with polym core order=3
SVM contr err with polym core order=3
SVM regrated error with polyn core order=4
N=5 SVM test error with polynomial core 4
Decision PCA rule. Number of classifier NumMod=
Regraded classif error of training samples

ErrRegrPCA =

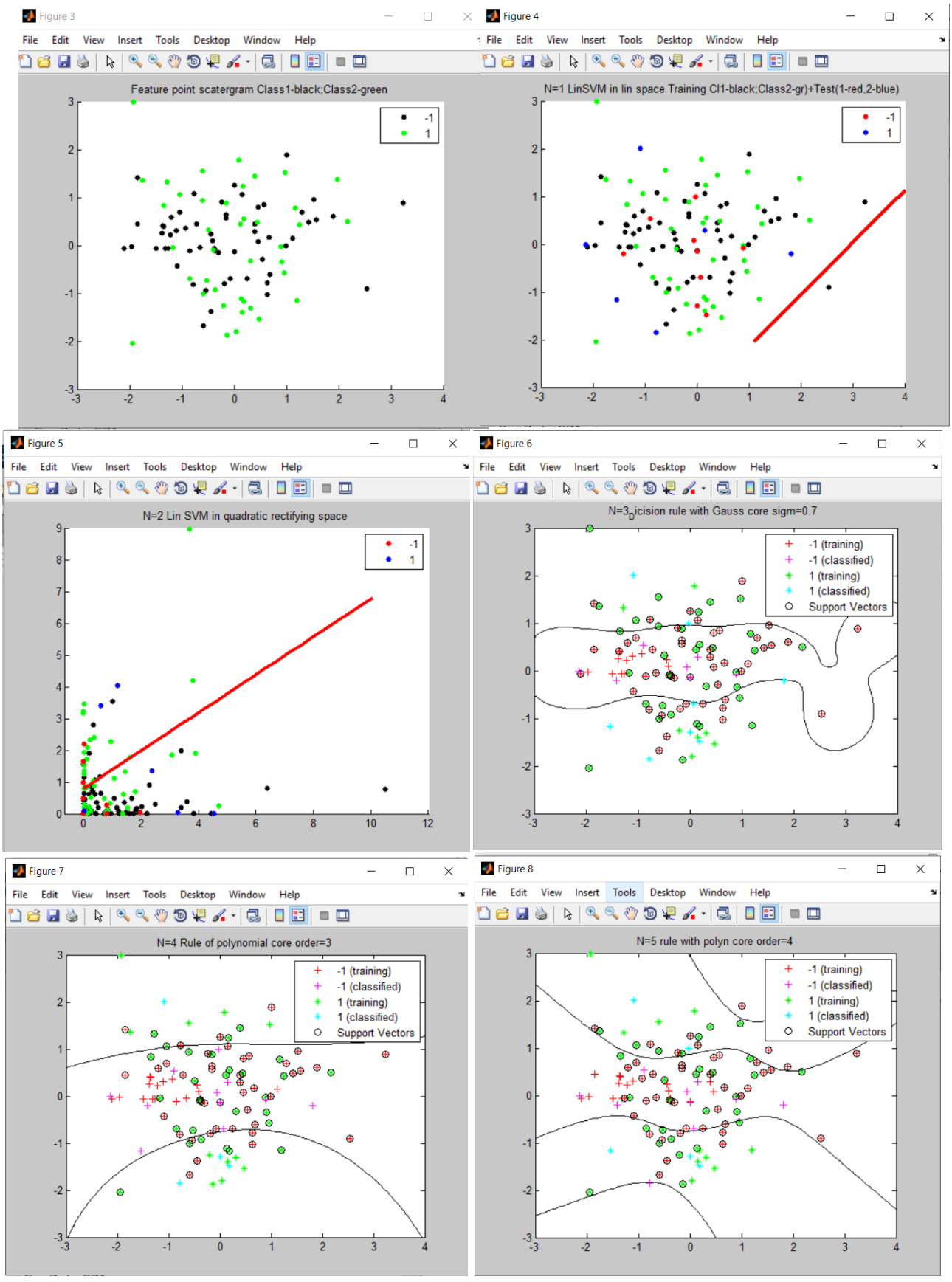
0.2050
0.1500
0.1400
0.1400
0.1400

Classif error of control samples

ErrCntrPCA =

0.4000
0.4667
0.4000
0.4000
0.4667

Analyse and save PCA results
Enter PCA rule number LinSVM in Linsp=1
lin SVM in quad space=2; Gauss core=3; Polyn core=4; Polyn core=5
NumModPCA=2



Analysis in the space of tuple of 3 and 1 features.

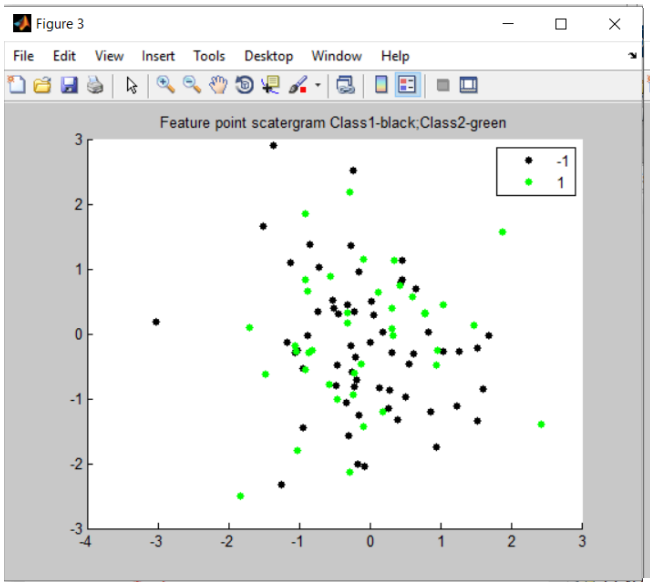
ErrRegrinf =

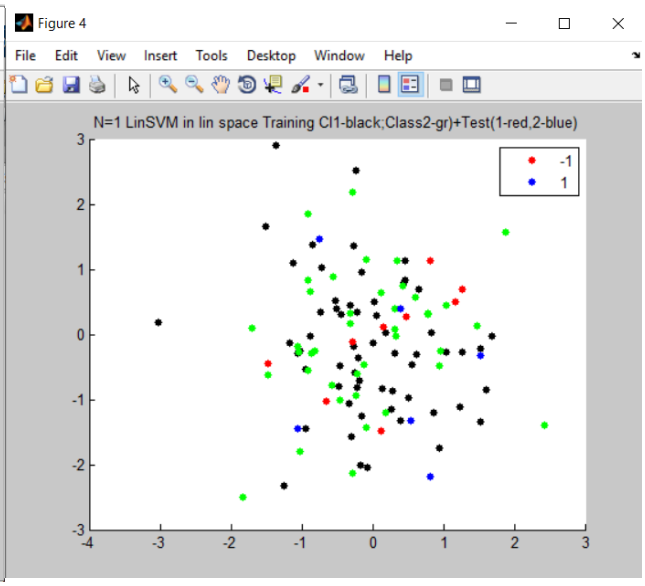
0.0618
0.1500
0.1400
0.1400
0.1400

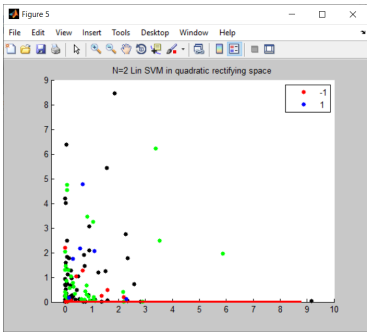
Classif error of test samples

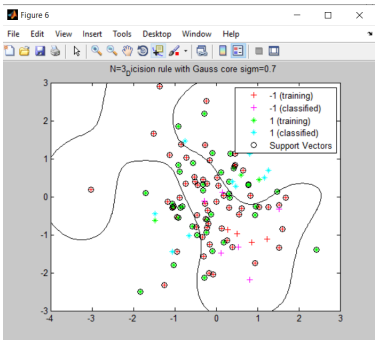
ErrCntrinf =

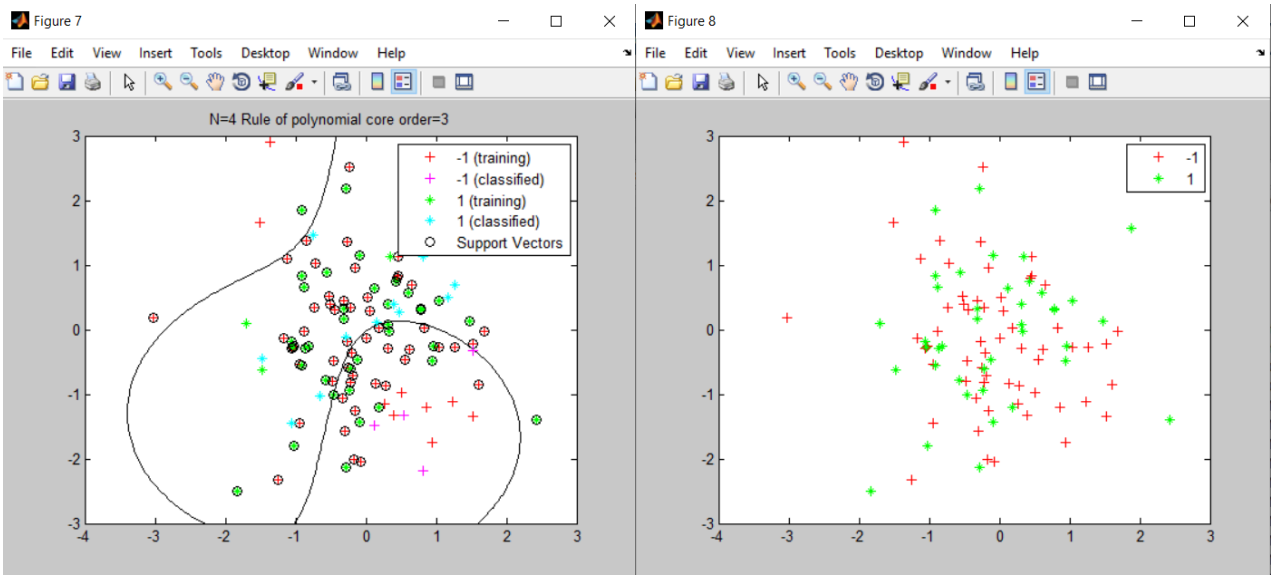
0.4000
0.4667
0.4000
0.4000
0.4667










Figure 7
File Edit View Insert Tools Desktop Window Help
N=4 Rule of polynomial core order=3
-1 (training)
-1 (classified)
1 (training)
1 (classified)
Support Vectors
Figure 8
File Edit View Insert Tools Desktop Window Help
-1
1