Algo Predictor - Prediction Software

AlgoExe Marcin Orchel

Abstract

Algo Predictor is a program/component for prediction based on support vector machines - one of the best machine learning methods in solving classification and regression problems. We implemented various quality and speed improvements, some of them are described on AlgoPre - research in predictive analytics site. We customized our product for quantitative analysis based on fundamental and technical factors. We offer two versions of Algo Predictor: basic and full. The basic version includes standard prediction methods. It is available for free on AlgoPre - research in predictive analytics site. The full version includes all implemented features and improvements.

Contents

Chapter 1

Introduction

1.1 Requirements

1. 64 bit operating system
2. 6 cores processor
3. 8 GB RAM
4. 1GB disk space
5. Java 7 64 bit installed

1.2 Installation

Download "algoPredictor.zip" from AlgoPre site and unpack all with the directory structure preserved. In "AlgoPredictor" directory, there is a file "algoPredictor.bat". You need to edit it and correct the path to your Java bin directory, and optionally the Xmx parameter.

1.3 Running

From the command line execute the program "algoPredictor.bat" which has 1 parameter, a path to the test. A test is a special file with the extension "test" which defines all parameters of the prediction.

1.4 Description of a Test File

All test parameters are set in a file with an extension "test". This is a java property file. We list all available parameters in Table 1.1, Table 1.2, Table 1.3, Table 1.4, Table 1.5.

Table 1.1: Parameters related to SVM

name	type	comment
unitTest	String	
solver	ESolverType	Possible values: advancedSVC, advancedFSVR, advancedESVR, advancedESVRSumSVR, advancedDeltaSVRSumSVR
kernel	String	
globalC	double	
isRegression	boolean	
isMulticlass	boolean	
isDistribution	boolean	
numberOfClasses	int	
isExtendedMatrix	boolean	
alphasLog	String	
slackVariablesLog	String	
possibleDetractorsLog	String	
functionalMarginsLog	String	
detailLog	String	
function	String	
dataCompression	double	
addConstraint	boolean	
valueConstraintLessOrEqual	double	
valueConstraintGreaterOrEqual	double	
gridSearchType	String	
doubleGridSearch	boolean	
maxIteration	int	

Table 1.2: Parameters related to advanced solvers of SVM

name	type	comment
heuristic	String	
subproblemSolver	String	
stopper	String	
functionDifferenceMinimum	double	

Table 1.3: Parameters related to δ-SVR

name	type	comment
translationParameterUp	double	
translationParameterDown	double	

Table 1.4: Parameters related to incorporating an additional constraint in the form of sum of values

name	type	comment
indices	String	

Table 1.5: All general test parameters

name	type	comment
testName	String	
testRef	String	
additionalVector	String	
useDetractors	boolean	
combinedData	int	
dualGapError	double	
checkDualGap	boolean	
generateNewData	boolean	
generateNewDataWithDetractors	boolean	
generateRandomDataWithDetractors	boolean	
generateDetractors	boolean	
automaticSVRemoval	int	
automaticSVRemovalWithDetractors	int	
combineWithDetractors	boolean	
checkTrainingMatrix	boolean	
assertTrainingMatrix	boolean	
assertTestingMatrix	boolean	
vector	String	
testVector	String	
trainingSetSize	int	
dataPosition	String	

Table 1.5: All general test parameters, cont.

testingSetSize	int
dataPath	String
testDataPath	String
outputTestDataPath	String
dataGenerator	String
generatorCount	int
weightsPath	String
weights	String
alphas	String
sameResults	boolean
alternativeValues	String
testingAlternativeValues	String
outputPath	String
shiftData	boolean
orderedData	boolean
automaticPercentageDetractors	boolean
writeCurve	boolean
writeGnuplot	boolean
separateCurves	boolean
curve	String
plotCurve	boolean
originalMethodCurve	boolean
plotOriginalResults	boolean
plotMargin	boolean
plotPoints	boolean
plotSV	boolean
plotGeneratedFrom	boolean
plotDensity	boolean
preserveAspectRatio	boolean
pointColor	String
pointType	String
lineColor	String
lineStyle	String
margin1	–
margin2	–
plotFormat	String
writeData	boolean
crossValidation	boolean
internalCrossValidation	boolean
cacheExtension	String
storeResult	boolean
multipleResult	boolean

Table 1.5: All general test parameters, cont.

storeResultPerDataFile	boolean
runAlone	boolean
showBestUnitTestLog	boolean

1.5 Examples of Test Files

1.5.1 support vector classification example

testName=support vector classification example
testDescription=
vector1=ci=10.0|phi=0.0|class=1|d=1 0
vector2=ci=10.0|phi=0.0|class=1|d=5 4
vector3=ci=10.0|phi=0.0|class=1|d=3 2
vector4=ci=10.0|phi=0.0|class=1|d=3 1
vector5=ci=10.0|phi=0.0|class=1|d=6 5
vector6=ci=10.0|phi=0.0|class=-1|d=0 1
vector7=ci=10.0|phi=0.0|class=-1|d=0 2
vector8=ci=10.0|phi=0.0|class=-1|d=0 4
vector9=ci=10.0|phi=0.0|class=-1|d=3 4
vector10=ci=10.0|phi=0.0|class=-1|d=5 6
vector11=ci=10.0|phi=0.0|class=-1|d=-3 0

heuristic=default
subproblemSolver=twoParameters
stopper=KKT
sameResults=true
alphas=
isRegression=false
isMulticlass=false
isDistribution=false
isExtendedMatrix=false
checkDualGap=false
functionDifferenceMinimum=0.0000000001
assertTrainingMatrix=false
assertTestingMatrix=false
trainingSetSize=
gridSearchType=errors
writeGnuplot=true
curve1

curve1.plotCurve=true
curve1.plotOriginalResults=false
curve1.plotMargin=true
curve1.plotPoints=true
curve1.plotSV=true
curve1.plotGeneratedFrom=false
curve1.lineColor=black
curve1.lineStyle=solid
curve1.margin1.lineColor=red
curve1.margin1.lineStyle=dashed
curve1.margin2.lineColor=green
curve1.margin2.lineStyle=dashed
plotFormat=LATEX
writeData=false
storeResult=false

unitTest1a
unitTest1a.solver=advancedSVC
unitTest1a.kernel=denseLinear 0.0
unitTest1a.globalC=100
unitTest1a.heuristic=default
unitTest1a.subproblemSolver=twoParameters
unitTest1a.stopper=PRIMARY_KKT
unitTest1a.slackVariablesLog=NONE
unitTest1a.possibleDetractorsLog=NONE
unitTest1a.alphasLog=NONE
unitTest1a.functionalMarginsLog=NONE
unitTest1a.detailLog=NONE
unitTest1a.showBestUnitTestLog=false

1.5.2 support vector classification example for random data

testName=support vector classification example for random data
dataGenerator=density|100|2|0.01|normal 0.3 0.7 0.02 0.0 0.0 0.02,normal 0.7 0.3 0.02 0.0 0.0 0.02
generatorCount=1
heuristic=default
subproblemSolver=twoParameters
stopper=KKT
sameResults=true
alphas=
isRegression=false
isMulticlass=false

isDistribution=false
isExtendedMatrix=false
checkDualGap=false
functionDifferenceMinimum=0.0000000001
assertTrainingMatrix=false
assertTestingMatrix=false
trainingSetSize=
gridSearchType=errors
writeGnuplot=true
curve1
curve1.plotCurve=true
curve1.plotOriginalResults=false
curve1.plotMargin=false
curve1.plotPoints=true
curve1.plotSV=false
curve1.plotGeneratedFrom=false
curve1.lineColor=black
curve1.lineStyle=solid
curve1.margin1.lineColor=red
curve1.margin1.lineStyle=dashed
curve1.margin2.lineColor=green
curve1.margin2.lineStyle=dashed
plotFormat=LATEX
writeData=false
storeResult=false

unitTest1a
unitTest1a.solver=advancedSVC
unitTest1a.kernel=denseLinear 0.0
unitTest1a.globalC=100
unitTest1a.heuristic=default
unitTest1a.subproblemSolver=twoParameters
unitTest1a.stopper=PRIMARY_KKT
unitTest1a.slackVariablesLog=NONE
unitTest1a.possibleDetractorsLog=NONE
unitTest1a.alphasLog=NONE
unitTest1a.functionalMarginsLog=NONE
unitTest1a.detailLog=NONE
unitTest1a.showBestUnitTestLog=false

1.5.3 support vector regression example

testName=support vector regression example
vector1=value=0|eU=0.5|ci=10.0|phi=0.0|d=1
vector2=value=5|eU=0.5|ci=10.0|phi=0.0|d=5
vector3=value=2|eU=0.5|ci=10.0|phi=0.0|d=3
vector4=value=1|eU=0.5|ci=10.0|phi=0.0|d=3
vector5=value=5|eU=0.5|ci=10.0|phi=0.0|d=6
vector6=value=1|eU=0.5|ci=10.0|phi=0.0|d=0
vector7=value=2|eU=0.5|ci=10.0|phi=0.0|d=0
vector8=value=4|eU=0.5|ci=10.0|phi=0.0|d=0
vector9=value=4|eU=0.5|ci=10.0|phi=0.0|d=3
vector10=value=6|eU=0.5|ci=10.0|phi=0.0|d=5
vector11=value=0|eU=0.5|ci=10.0|phi=0.0|d=-3

heuristic=default
subproblemSolver=twoParameters
stopper=KKT
sameResults=true
alphas=
isRegression=true
isMulticlass=false
isDistribution=false
isExtendedMatrix=false
checkDualGap=false
functionDifferenceMinimum=0.00000001
assertTrainingMatrix=false
assertTestingMatrix=false
trainingSetSize=
gridSearchType=errors
writeGnuplot=true
curve1
curve1.plotCurve=true
curve1.plotOriginalResults=false
curve1.plotMargin=true
curve1.plotPoints=true
curve1.pointColor=black
curve1.pointType=plus
curve1.plotSV=true
curve1.plotGeneratedFrom=false
curve1.lineColor=black
curve1.lineStyle=solid
curve1.margin1.lineColor=black
curve1.margin1.lineStyle=dashed

curve1.margin2.lineColor=black
curve1.margin2.lineStyle=dashed
curve2
curve2.plotCurve=false
curve2.plotOriginalResults=false
curve2.plotMargin=false
curve2.plotPoints=false
curve2.plotSV=false
curve2.plotGeneratedFrom=false
curve2.lineColor=black
curve2.lineStyle=solid
curve2.margin1.lineColor=black
curve2.margin1.lineStyle=dashed
curve2.margin2.lineColor=black
curve2.margin2.lineStyle=dashed
plotFormat=LATEX
writeData=false
storeResult=false

unitTest1a
unitTest1a.solver=advancedESVR
unitTest1a.kernel=denseLinear 0.0
unitTest1a.globalC=100
unitTest1a.epsilonParameter=0.5
unitTest1a.heuristic=default
unitTest1a.subproblemSolver=twoParameters
unitTest1a.stopper=PRIMARY_KKT
unitTest1a.slackVariablesLog=NONE
unitTest1a.possibleDetractorsLog=NONE
unitTest1a.alphasLog=NONE
unitTest1a.functionalMarginsLog=NONE
unitTest1a.detailLog=NONE
unitTest1a.showBestUnitTestLog=false

1.5.4 support vector regression example for random data

testName=support vector regression example for random data
dataGenerator=Y|100|1|0.06|Linear 1.0 0.0
generatorCount=1
heuristic=default
subproblemSolver=twoParameters
stopper=KKT
sameResults=true

alphas=
isRegression=true
isMulticlass=false
isDistribution=false
isExtendedMatrix=false
checkDualGap=false
functionDifferenceMinimum=0.00000001
assertTrainingMatrix=false
assertTestingMatrix=false
trainingSetSize=
gridSearchType=errors
writeGnuplot=true
curve1
curve1.plotCurve=true
curve1.plotOriginalResults=false
curve1.plotMargin=false
curve1.plotPoints=true
curve1.pointColor=black
curve1.pointType=plus
curve1.plotSV=false
curve1.plotGeneratedFrom=false
curve1.lineColor=black
curve1.lineStyle=solid
curve1.margin1.lineColor=black
curve1.margin1.lineStyle=dashed
curve1.margin2.lineColor=black
curve1.margin2.lineStyle=dashed
curve2
curve2.plotCurve=false
curve2.plotOriginalResults=false
curve2.plotMargin=false
curve2.plotPoints=false
curve2.plotSV=false
curve2.plotGeneratedFrom=false
curve2.lineColor=black
curve2.lineStyle=solid
curve2.margin1.lineColor=black
curve2.margin1.lineStyle=dashed
curve2.margin2.lineColor=black
curve2.margin2.lineStyle=dashed
plotFormat=LATEX
writeData=false
storeResult=false

unitTest1a
unitTest1a.solver=advancedESVR
unitTest1a.kernel=denseLinear 0.0
unitTest1a.globalC=100
unitTest1a.epsilonParameter=0.1
unitTest1a.heuristic=default
unitTest1a.subproblemSolver=twoParameters
unitTest1a.stopper=PRIMARY_KKT
unitTest1a.slackVariablesLog=NONE
unitTest1a.possibleDetractorsLog=NONE
unitTest1a.alphasLog=NONE
unitTest1a.functionalMarginsLog=NONE
unitTest1a.detailLog=NONE
unitTest1a.showBestUnitTestLog=false

Please download more tests and data sets from AlgoPre site.