IMT-4612
Gjovik Univiristy Collage, Spring 2014

Author: Ole Henrik Paulsen
Student number: 130572

Assigment 1

This document was compiled on February 17, 2014

Contents

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam quis pellentesque nibh. Curabitur rhoncus non risus sit amet pharetra. Praesent nisi libero, ultrices sed ultricies vitae, hendrerit id lectus. Nulla facilisi. Sed in adipiscing magna, at faucibus arcu. Mauris tristique consectetur diam, ultrices semper purus ornare vel. Cras posuere tempor pharetra. Donec suscipit convallis congue. In faucibus tempus lacus at viverra. Nulla facilisis felis non odio rutrum placerat. In hac habitasse platea dictumst. Donec tincidunt mauris augue, eget auctor tortor consectetur placerat. Sed at est et mauris congue egestas eu at lectus. Vivamus suscipit blandit iaculis.

1 Introduction

The first assigment in Machine Learning and Pattern Recognition 1.

2 Data Analysis and Knowledge Representation

2.1 Cartesian coordinate system

Picture comming here This is done first in matlab with the function gscatter(X,Y,Z) Since my lack of matlab skills I did play with Python and the matplotlib libary to make scripts that show the same thing to understand it better. The Python script:

Python script is comming here

Resultetd in mostly the same Coordinate system: *Picture is comming here*

Possibilties to applye linear classification models..

There is a possibility to apply a Classification here. We have 2 groups here. Label 1 and Label 2. Label 1 have a y max value on 8, and a X max value on 9. Label 2 have a Y min on 9 and a X min on 11. So here can we split the Cordinate system with two lines. One line horizontial where Y= 8,5 and one vertical where X= 10.

Potential uncertainties in classification due to overlapping of samples in different classes... This group of cords are close so there can be some uncertanties. An other argument that can apply to more uncertanties are that there is small amount of train data. The statement about small amount of train-data can you easy see when you plot the test data, where some of the cords it's outside the rules based on the traning data.

2.2 Boundaaries

Picture of test data in coordinate

2.3 Areaplot

First i did build the Areaplot in Matlab, but here aswell i scriptet a Python Script with the libary matplotlib to understand more of the graph.

Picture of matlab picture

Python script

Picture of python result

It's easy to see the two groups here on the vizualization. You can easy see the boundaaries on the Y-axis where the values are lower.

2.4 Weka 1

First thing i did was to convert the txt files with test and training data into arff files (train.arff and test.arff)

Loaded the files into Weka, Using the Attribute Evaluator: ClassifierSubsetEval and Search Method: GreedyStepwise

The output is following:

```
=== Run information ===

Evaluator:    weka.attributeSelection.ClassifierSubsetEval -B weka.classifiers.rules.ZeroR -T -H
Search:weka.attributeSelection.GreedyStepwise -T -1.7976931348623157E308 -N -1
Relation:     lol2
Instances:    5
Attributes:   2
              var1
              var2

Evaluation mode:evaluate on all training data

=== Attribute Selection on all input data ===
```

```
Search Method:
Greedy Stepwise (forwards).
Start set: no attributes
Merit of best subset found:    4.08

Attributesute Subset Evaluator (supervised, Class (numeric): 1 var1):
ClassifierSubsetEval Subset Evaluator
Learning scheme: weka.classifiers.rules.ZeroR
Scheme options:
Hold out/test set: Training data
Accuracy estimation: MAE

Selected attributes:
```

```
=== Run information ===

Evaluator:    weka.attributeSelection.ClassifierSubsetEval -B weka.classifiers.rules.ZeroR -T -H
Search:weka.attributeSelection.GreedyStepwise -T -1.7976931348623157E308 -N -1
Relation:     lol2
Instances:    5
Attributes:   2
              var1
              var2
Evaluation mode:evaluate on all training data

=== Attribute Selection on all input data ===
Search Method:
Greedy Stepwise (forwards).
Start set:      no attributes
Merit of best subset found:    3.44

Attribute Subset Evaluationvaluator (supervised, Class (numeric): 2 var2):
Classifier Subset Evaluationvaluatoruator
Learning scheme: weka.classifiers.rules.ZeroR
Scheme options:
Hold out/test set: Training data
Accuracy estimation: MAE

Selected Attributettributes:
```

2.5 Weka 2

3 Machine learning

3.1 Computer program

3.2 Techniques applied to increase generalization

3.3 Complexity

3.4 Metrics

References

[1] Wikipedia. Lorem ipsum — wikipedia, the free encyclopedia, 2013.

[Online; accessed 20-October-2013].