Machine Learning and Pattern Recognition — IMT4612
Gjøvik University Collage, Spring 2014

Author: Ole Henrik Paulsen
Student number: 130572

Assigment 2

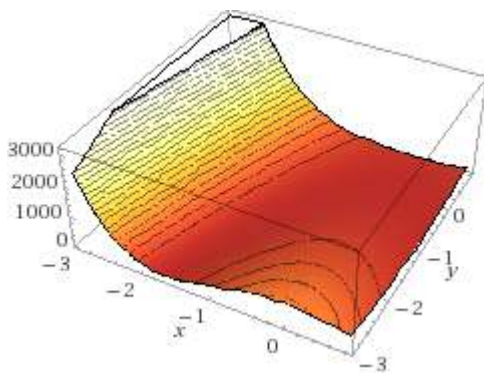# Contents

Abstract

Assigment 2 in Machinelearning

1 Learning as a Search

1.1 Global optimal solutions

The global minima in the fitness function are (-1,-1)

This can be calculated with this equaction:

min(75(y + x**2)**2 + (1 +x)**2) = 0

1.2 Genetic Algorithm(GA)

This section is done in Python with the library Numpy for max, min and array handling(matrix).

I made a Class that takes alot of variables that you easy can edit to optimize the genetic algorithm.

The seed is made of random numbers between two values. random(-10,10) will give you a random number from -10 to 10.

The selection function are selecting the 4 best chromosomes and deleting the rest of the set.

Crossover are switching randomly x-values or y-values from a perrent with an other perrent. The new pair are added to the set for next generation.

There is also spawning a new random pair of (X,Y) evry generation. The only ruule for this new random spawned pair is that is not alleready in the set of cromosomes.

The fitnes function is takeing a X and Y value as input and returing the number of the equation.

When the final generation of rouds are done it hopyfully retturing the cords -1,-1.

1.3 Gradient Descent method

1.4 Performance domain

2 Statistical Learning

2.1 Computer program

I did the programming in Python with the library Numpy and Time. You need to install Numpy, but time is a core library of Python. Numpy are used to handle sqrt, max, min and mathematical functions on arrays. I also used Numpy to read in data from the txt files. Numpy also have functions for Euclidean and chebyshev, but they are NOT used in my program.

K Neaarest Neighbor are also programmed by from the bottom instead of using a library.

K nearst Neighbor code

The program are scaled to handle large amount of input, both train and validation data, and 13 attributes takes under 3 secounds to handle.

2.2 Read input files

The train.txt and validation.txt are read into the program with Numpy's genfromtxt. *genfromtxt code*

I save the input data in a masked arrays. and split out the label into a own array.
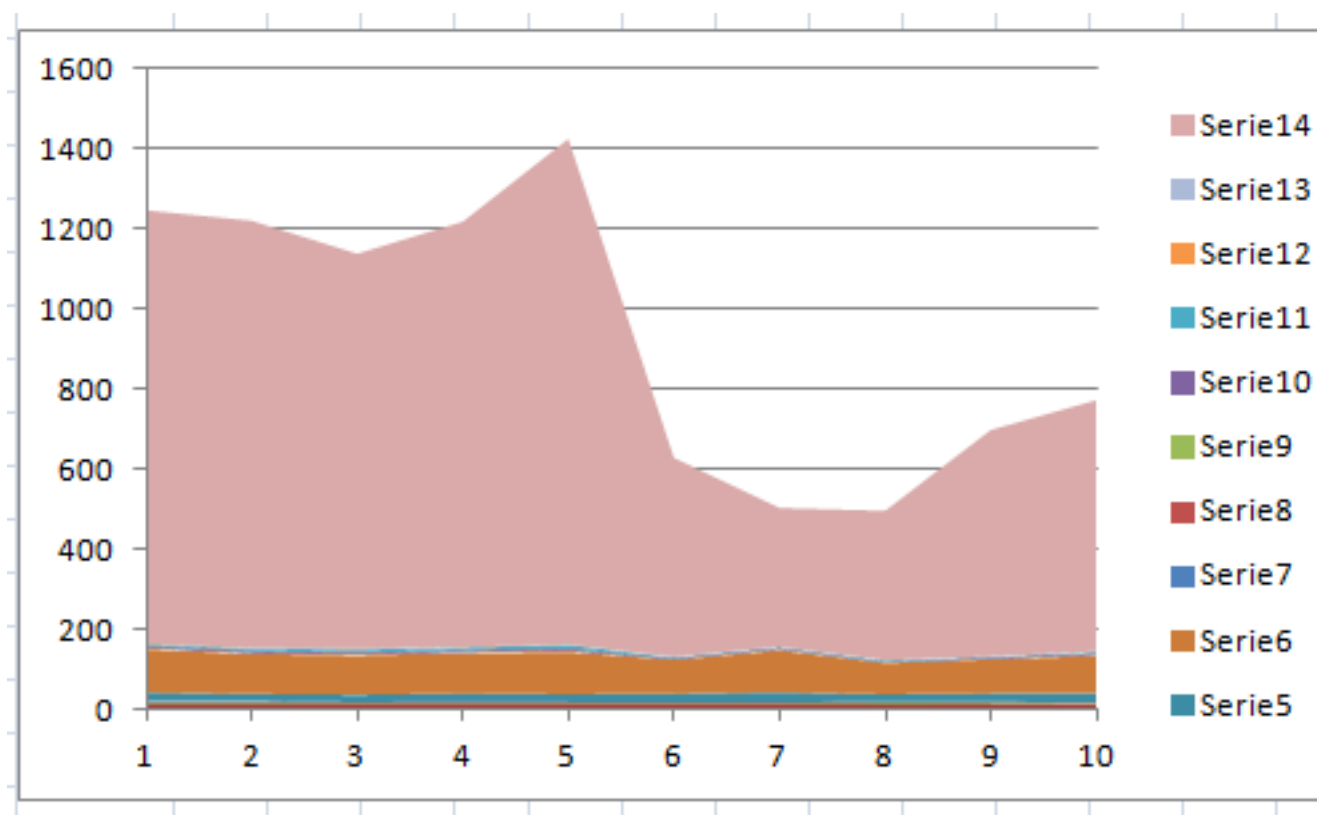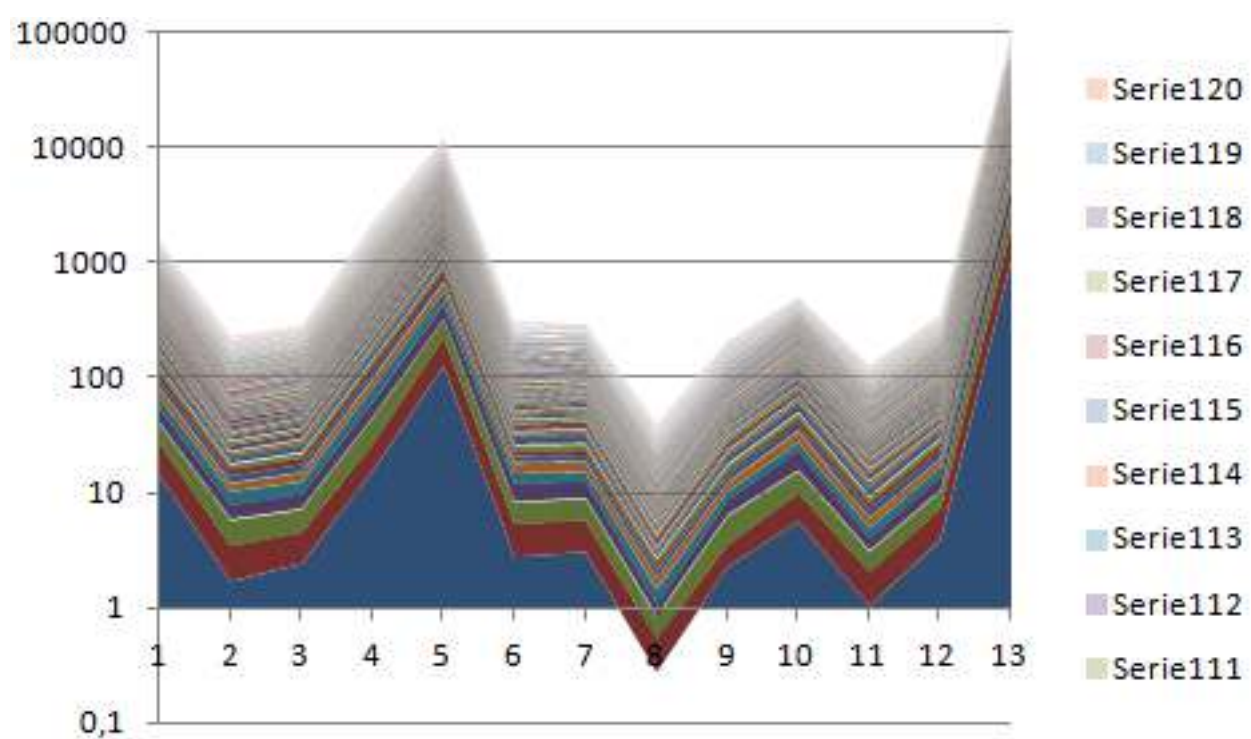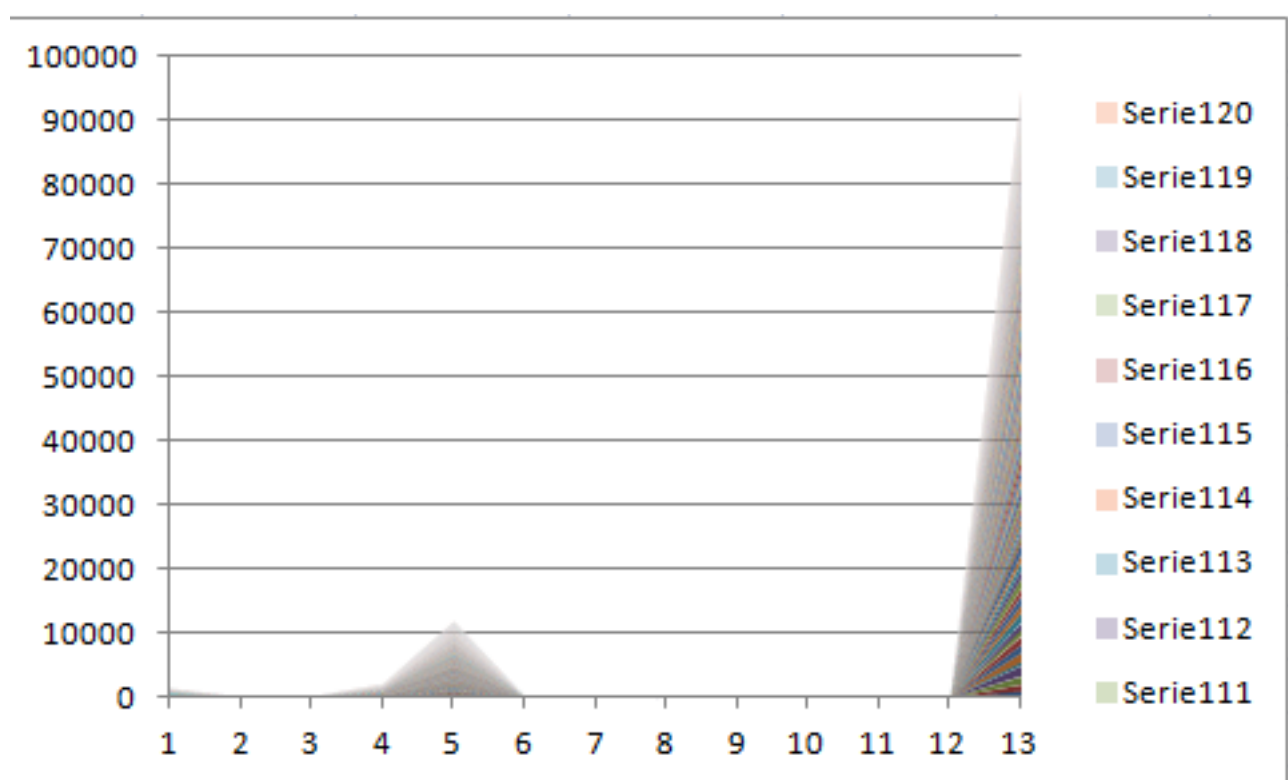
Splitting code

I confirm the input is 120 train samples and 10 validation samples in the output.
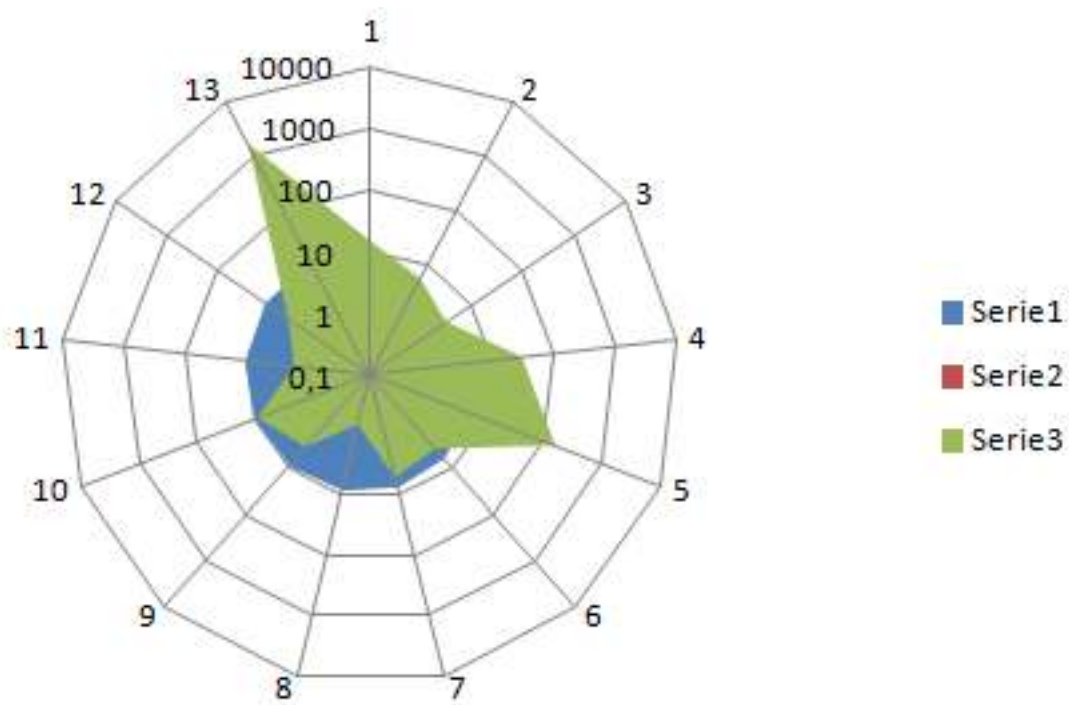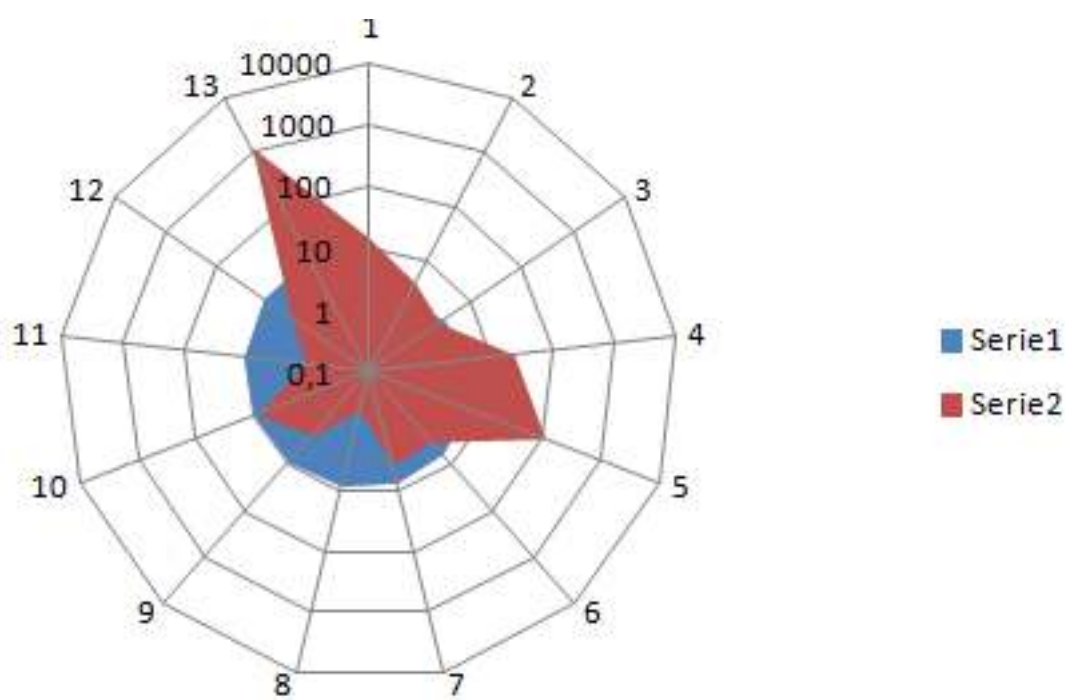
2.3   Radar and Area plot

I built the graphs under in Excel afterr hours of trying and failing in matplotlib and numpy.

Some of the graphs are using logarithmcs scala on the axis.

Serie120
Serie119
Serie118
Serie117
Serie116
Serie115
Serie114
Serie113
Serie112
Serie111

Serie120
Serie119
Serie118
Serie117
Serie116
Serie115
Serie114
Serie113
Serie112
Serie111

*

2.4 Distance Algorythms

The follow three algorithms are included into the program:

Euclidean:

Squar Euclidean:

Chebyshev:

As you can see they got different output from each other

2.5 Output

The output of the program is as follow:

Output from the program

References

[1] WIKIPEDIA. Lorem ipsum — wikipedia, the free encyclopedia, 2013. [Online; accessed 20-October-2013].