Assignment 3B

David Boerema (s3683869)
Marios Souroulla (s4765125)
Marius Captari (s4865928)
Max Valk (s3246922)

Group 3

October 12, 2021

3B1

For this assignment, the script `myGeneticAlgorithm.m` has the following functions filled out:
- `generateInitialPopulation`: Generate a n by ndim array of random binary values
- `getOffSpring`: Perform single point crossover with the crossover point k in the range $[2, ndim\}$. Afterwards perform mutation by generating $ndim$ random values between 0 and 1 and then flipping the bit if it's below the mutation probability.
- `getScore`: Calculate the score using $score = (10^4 \times accuracy) + (0.4 \times \#zeros)$. Please note that we calculate the amount of zeros by subtracting the sum of the chromosome (which gives us $\#ones$) from the total length of the chromosome.

3B2

The goal of this part was to use the the genetic algorithm implemented in the previous part to make feature selection and then use a KNN classification method to classify the WINE data-set and compare the classification accuracy with all features.

For this part, we run 15 independent experiments and compare the accuracy of the KNN classifier using all the features (13) versus the accuracy of the KNN classifier using the features selected by the myGeneticAlgorithm.m function. The number of features selected by the genetic algorithm vary from 3 to 7 (for this run).

The comparisons were made using the comparisons.m Matlab script.

As can be seen from the figure 1, the genetic algorithm may choose different chromosome in each run which means different features (and number of features) will be used in the KNN classifier, this is expected and it comes from the randomness that the genetic algorithm uses during the mutation phase and the random initial population.

Results

For our run of our experiment we got an average accuracy of 91.45% using the genetic algorithm method and an average accuracy of 74.96% using all the features. The minimum number of features selected by the genetic algorithm is 3 and the maximum number of features is 7.

It is notable that, in all 15 runs the accuracy of the genetic algorithm was far superior than the accuracy of all the features. Moreover, it reached an accuracy of 96%.

These results are just one run and each time we run the experiments different results will occur but these are the main conclusions which arise every time.
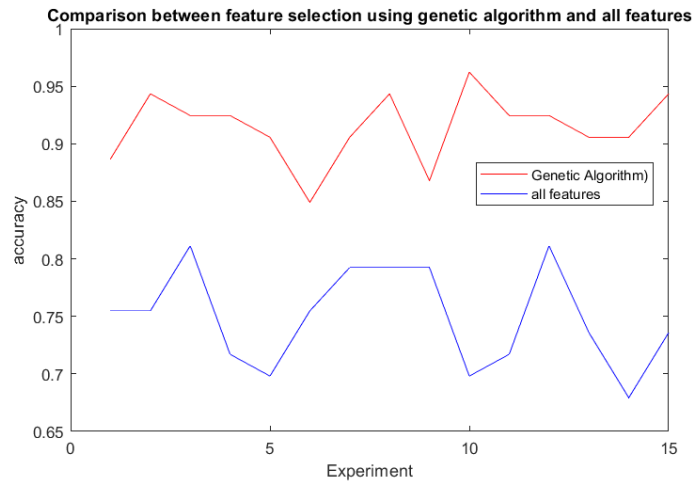
Figure 1: Accuracy of Genetic Algorithm vs all features

Figure 1 shows the accuracy of the Genetic algorithm vs the accuracy of all the features on the WINE dataset