159.372 Intelligent Machines: Assignment 1

Weight: 20%

Due: September 15, 2019

Late submissions will receive 3% penalty for each day that they are late. They will not be accepted at all if they are more than one week late.

1 Bank marketing

In this assignment we will use the bank marketing dataset from UCI machine learning repository to classify whether a given customer will subscribe to a banking product (term deposit) based on features extracted during direct marketing campaigns of a Portuguese banking institution. Download the dataset from repository and go through the description. For detailed insights about the dataset refer to these publications: [1] and [2]

2 Tasks

Implement these tasks in Python. Feel free to reuse the code implementation from the lab exercises.

- **Classification with Multi Layer Perceptron (30 marks)**
  - implement two class output encoding for the MLP;
  - perform any preprocessing of the data;
  - train a simple MLP (choose a number of hidden nodes that seems reasonable) and see how well you can perform the classification. Use the confusion matrix to output the results;
  - implement cross-validation;
  - test out different sizes of hidden layer to see how many hidden neurons give the best results;
- **Improve MLP using genetic algorithm (40 marks)**
  In the above task you are using standard back propagation algorithm to adjust the weights in the neural network. In this part you are required to optimize the weights in MLP using genetic algorithm.
  - encode the weights of a neural network in a string representation;
  - code a fitness function to evaluate the fitness of the chromosome representing the weights of the MLP

  - test out the algorithm and compare the classification accuracy with the MLP trained using back propagation

- **Classification using Self Organising Map (30 marks)**
  - Use the SOM to cluster the data and see whether you can identify the subscribed customers in the clusters;
  - Use a Perceptron to take the activations of the SOM neurons as input and learn the outputs classes. How well does this work compared to the MLP?

References

[1] S. Moro, P. C., and Rita., P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems 62*, 22–31.

[2] S. Moro, R. L., and Cortez, P. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of the European Simulation and Modelling Conference - ESM'2011*, P. N. et al., Ed., pp. 117–121.

3 Deliverables

A brief report highlighting what you did, together with your (commented) code. The report should include the following things:

- description of the MLP that you used, with the optimal parameters and results you got, plus mention what preprocessing, etc. you performed
- the parameters that the genetic algorithm selected for the MLP, and the results of that network
- some comments on the difference between the outputs of the two methods of training the MLP.
- the results of using the SOM and some comments on the difference in results by using that method
- any extra things that you did

Zip everything together and submit it via Stream.