

COSC 4P76 - Machine Learning

Project proposal by Dennis Ideler

A comparative study of at least two different classification techniques.

Train at least two classifiers for steel plates faults diagnosis [1] and handwritten digits recognition [2]. The first dataset contains 27 attributes, has 7 classes, and 1941 instances. The second dataset contains 256 attributes, has 9 classes, and 1593 instances. Both datasets do not contain any missing values.

Classifiers to be used are a multi-layer neural network with backpropagation and k-nearest neighbour. Variations such as quickprop or weighted *k*-NN will be included if time permits.

Classifiers will be fine-tuned and the results will be (statistically) analyzed to determine which classifier performed "best" on this given dataset. Feature selection techniques will be used to try and improve the accuracy, along with data normalization.

Results and analysis will include: parameter tables, data visualization (e.g. graph plots), comparison of means, standard deviation, accuracy comparisons, and possibly accuracy vs time ratios (taking into account that running time is affected by implementation).

References to related material:

[1] [Steel plates faults dataset](#)

[2] [Semeion handwritten digits dataset](#)