Motivation and Contribution

• Understand how Extreme Learning Machines (ELM) work
• Identify strengths and weaknesses of ELM
• Combine ELM and Genetic Algorithm (GA) for faster feature selection
• Produce code to combine Cascade and ELM to create a new type of network (CASCELM)
• Identify strengths and weaknesses of CASCELM through experimentation
Extreme Learning Machines

- Uses the same topology as a multilayer perceptron
- Only the output layer is trained
- Pros: Very quick to train
- Cons: Performance is variable
- Similar to Echo State Machines


Comparison of ELM and Backprop

Original Matlab code from: Nanyang Technological University, Singapore

- **Back Propagation**
  - Hidden layer: 20 nodes
  - Training time: 6.4 S
  - Accuracy: 100%

- **ELM**
  - Hidden layer: 20 nodes
  - Training time: 0.02 S
  - Accuracy: 69%

- **ELM**
  - Hidden layer: 200 nodes
  - Training time: 0.066 S
  - Accuracy: 97%
Variations in training accuracy

25 nodes in hidden layer

50 nodes in hidden layer
The problem of over fitting with ELM

Output of diabetes training and test data from ELM
Cascade Correlation Networks

• A Constructive algorithm
• Pros:
  – Do not require back propagation
  – Produces more optimum topologies
• Cons:
  – Can lead to very deep networks which are slow to compute
• Cascade-Correlation: Scott Fahlman and Christian Lebiere, 1991
• Layered-CasPer was suggested by Tom Gedeon in 2011
Shallow ELM Cascade

Shallow cascade performs better than ELM with smaller hidden layers
Multiple ELM machines are cascaded but only the output from the preceding layer is cascaded to the next
In experiment 2 the output from all preceding layers are cascaded
Cascaded ELM results summary

- Over fitting is less evident with CascELM for the Diabetes training set
- Training times are longer (*10 for 250 hidden neurons)
Feature selection using ELM and genetic algorithms

During the project ELM code was provided to another student for research using genetic algorithms for feature selection

“This project is about producing a stress curve to display individuals' stress level by using their physiological signals.”

“Since it was estimated that such training may cost approximately 13 weeks, we decided to replace the back-propagation with ELM. Then just by one week, we finished the training.”

Xuanying Zhu
Conclusion and further work

ELM provide an efficient alternative to back-propagation when training Neural Networks

CASELM works but further research is needed:

Further research is needed for better random weight selection:
- Range of weights
- Distribution of weights

Further experiments with cascade:
- Incremental testing of each cascade
- Use of cascade to train on infinite data
- Iterative refinement of random layer weights