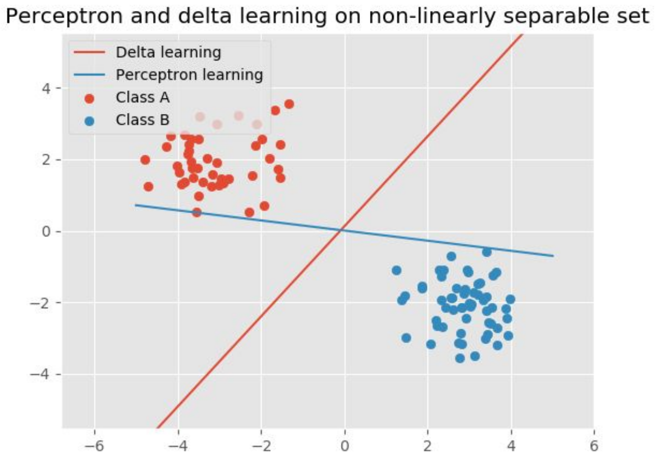
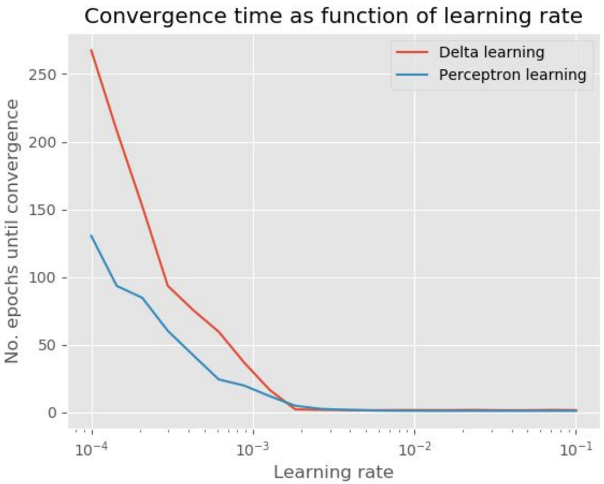
Group 15 - LAB 1

Mustafa Al-Janabi | Einar Lennelöv | Akanshu Mahajan

3.1
Single Layer Perceptron

Perceptron learning vs Delta rule

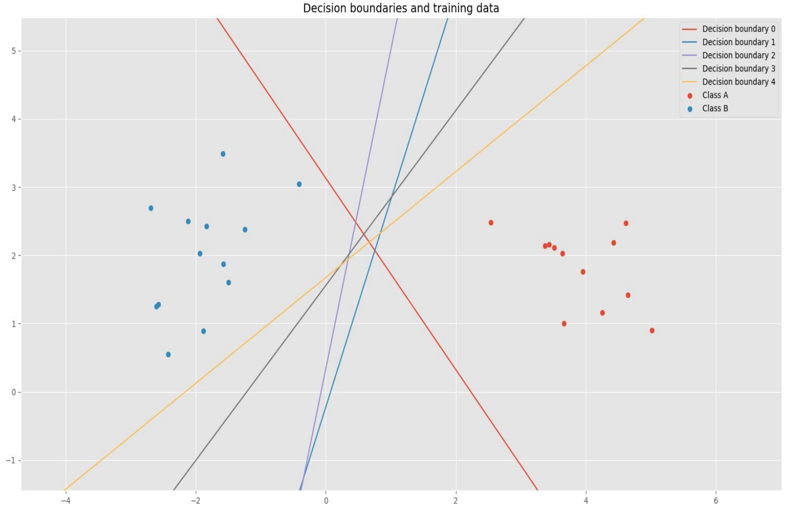


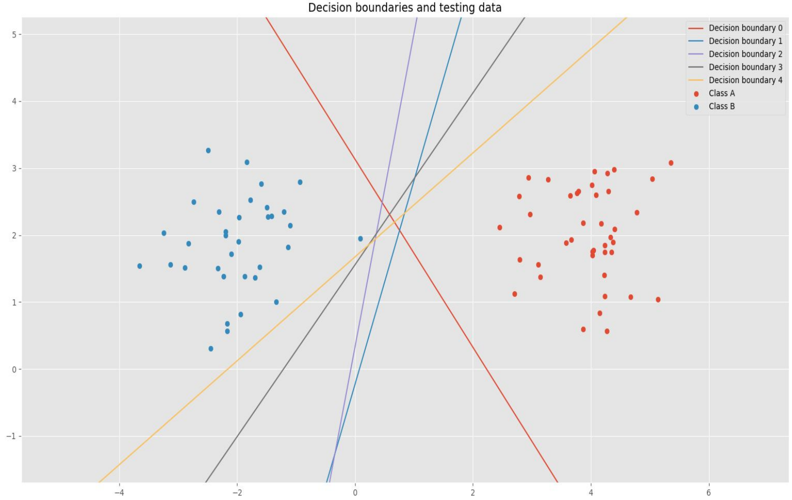


Comparison between delta learning and perceptron learning on a single dataset.

Delta rule

Sequential Learning

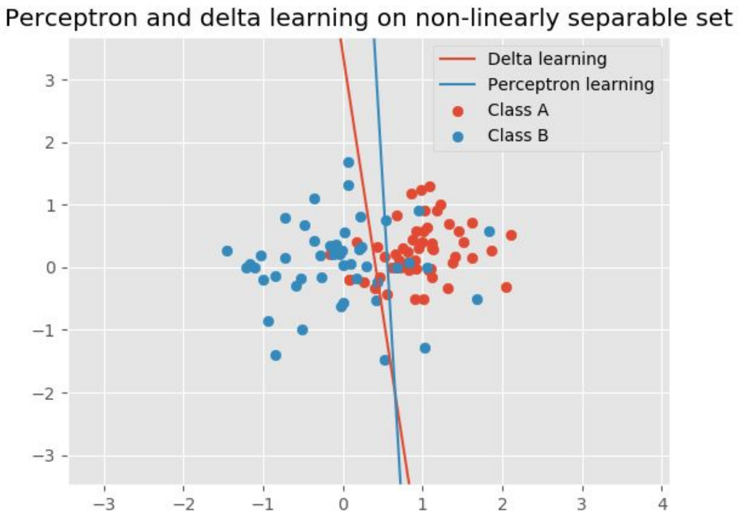


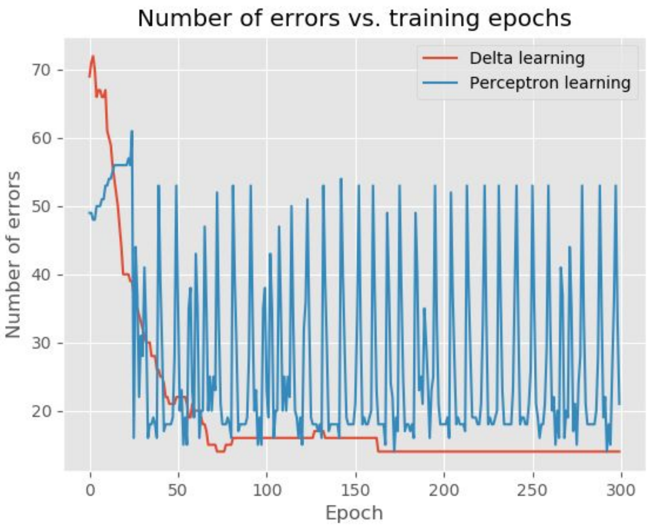


Delta learning rule applied to a single layer network using sequential learning rule instead of batch learning. (Epochs = 100 and learning rate = 0.001).

Conclusions: The network performed well in terms of classifying both the train data as well as the test data with an accuracy of 100%.

Non linearly separable





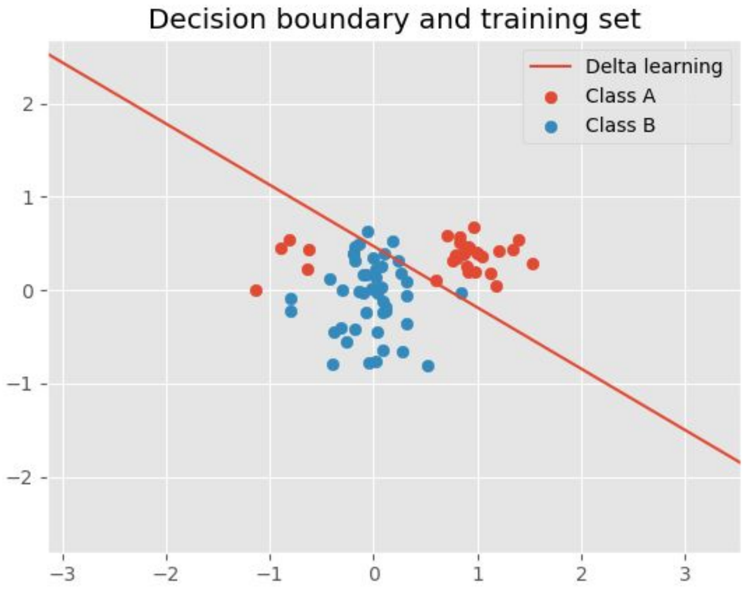
Comparison between perceptron learning and delta learning on linearly non-separable dataset.

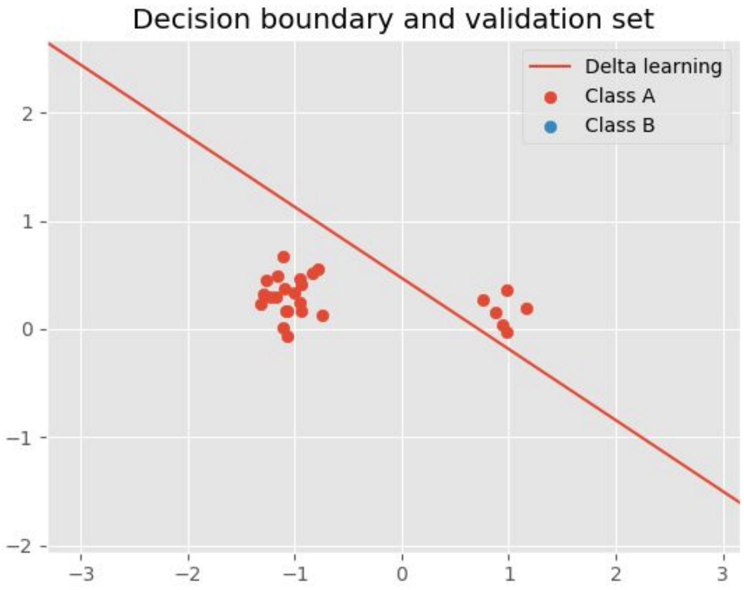
Subsampling

Part of A	Part of B	Mean accuracy* ± 1std.
25%	25%	0.76 ± 0.06
50%	0%	0.58 ± 0.09
0%	50%	0.43 ± 0.08
{20% A[0] > 0} U {80% A[0] < 0}	0%	0.23 ± 0.00

*Averaged over 100 runs.

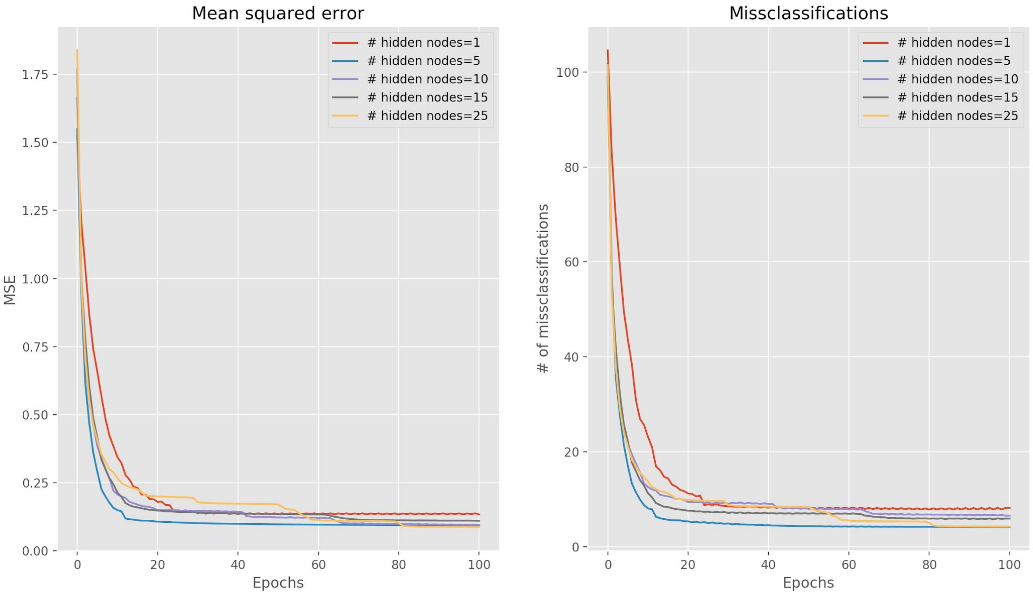
Subsampling





3.2
Multi layer perceptron

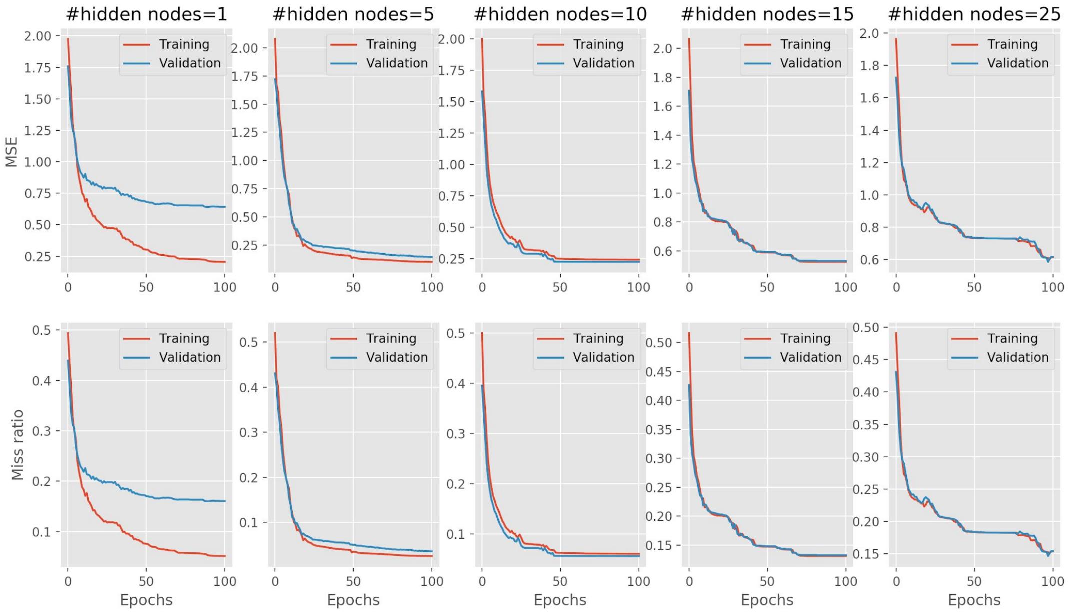
Classification of non linearly separable



Average over 100 runs for different number of nodes in the hidden layer.

Conclusions: Fastest convergence and best performance with 5 hidden nodes.

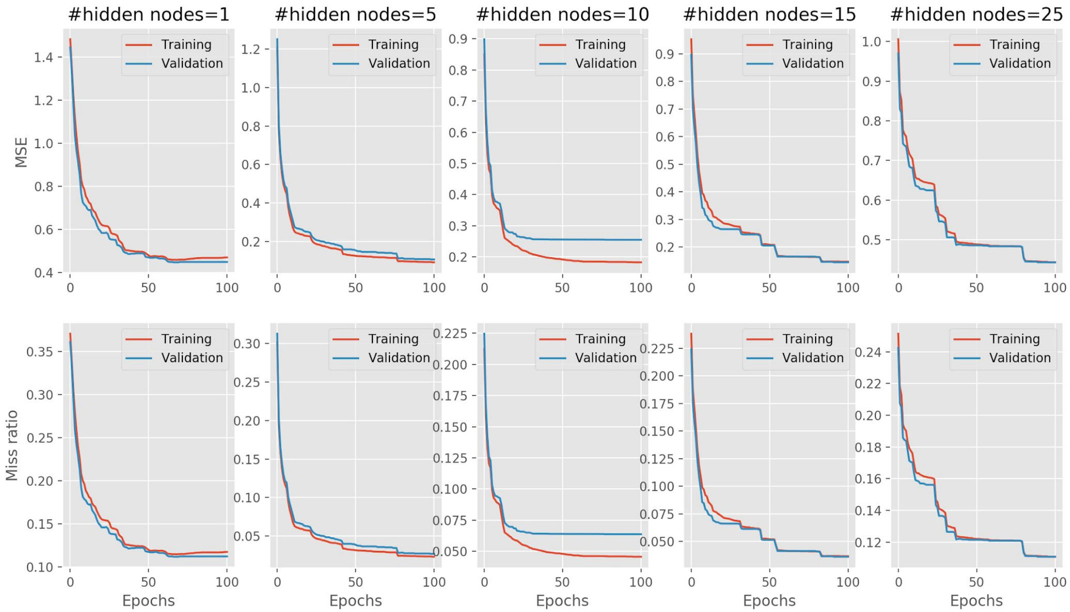
Generalisation



Average over 50 runs for different number of nodes in the hidden layer for training and validation sets.

Conclusions: Generalization is achieved for more than 10 hidden nodes.

Batch vs sequential

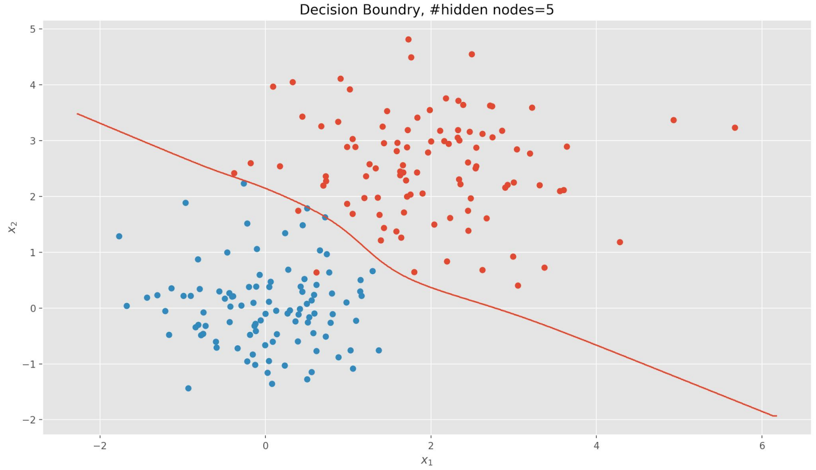


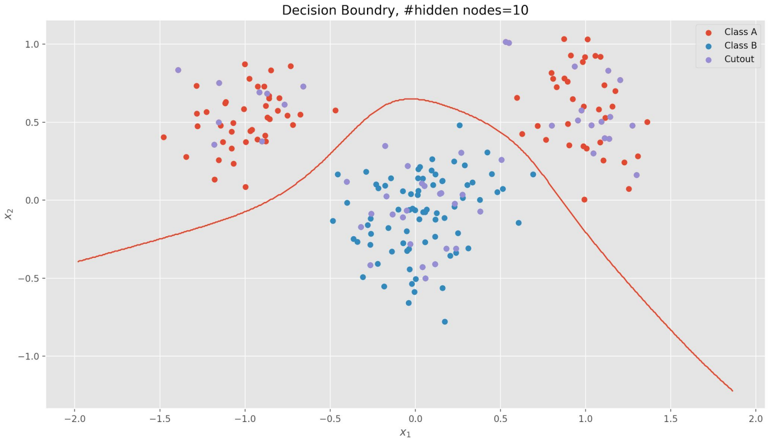
Same as previous slide, but using sequential learning.

Conclusions: Difficult to deduce differences.

Decision boundary

With 5 hidden nodes and learning rate η=0.01 we get

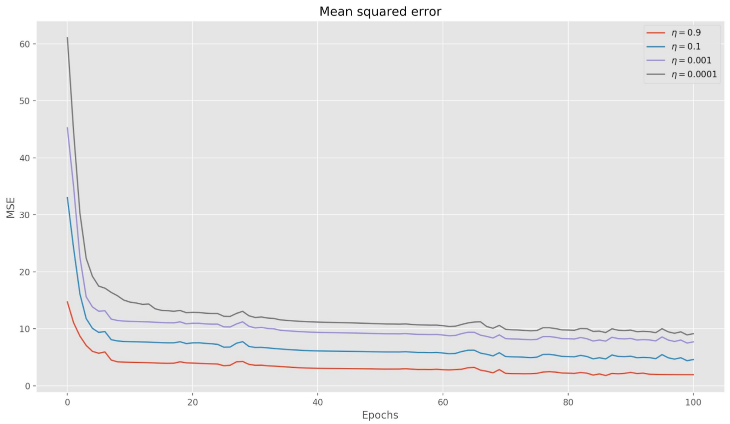




2 misclassifications with MSE=0.05 for the regular case

0 misclassifications with MSE=0 for training data and MSE=0 for the special case

Encoder problem 3 hidden layers



X = [-1, -1, 1, -1, -1, -1, -1, -1]
H = [-1,-1,-1]

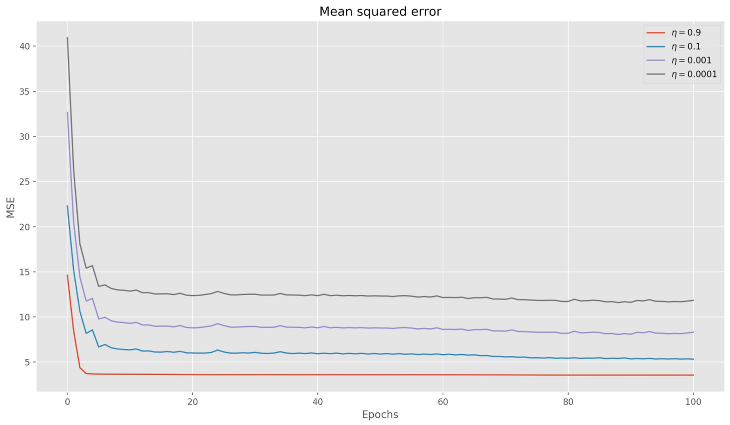
X = [-1, -1, -1, -1, -1, 1, -1, -1]
H = [1, 1, -1]

X = [1, -1, -1, -1, -1, -1, -1, -1]
H = [1, 1, 1]

Converges for high η.

Conclusions: Fastest convergence and best performance with 5 hidden nodes.

Encoder problem 2 hidden layers



X = [-1, -1, 1, -1, -1, -1, -1, -1]
H = [-1,-1]

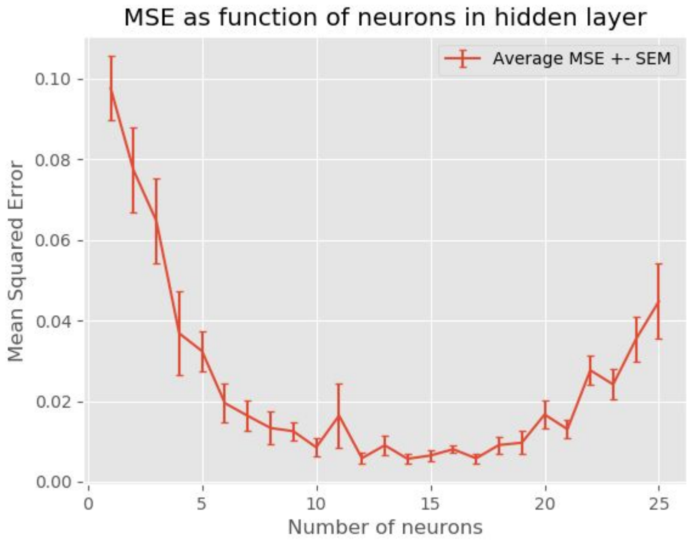
X = [-1, -1, -1, -1, -1, 1, -1, -1]
H = [-1, -1]

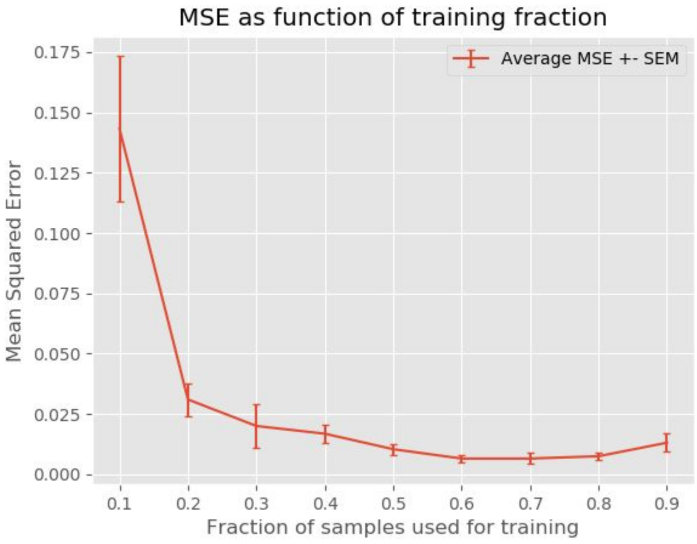
X = [1, -1, -1, -1, -1, -1, -1, -1]
H = [1, -1]

Converges for high η.

Conclusions: Fastest convergence and best performance with 5 hidden nodes.

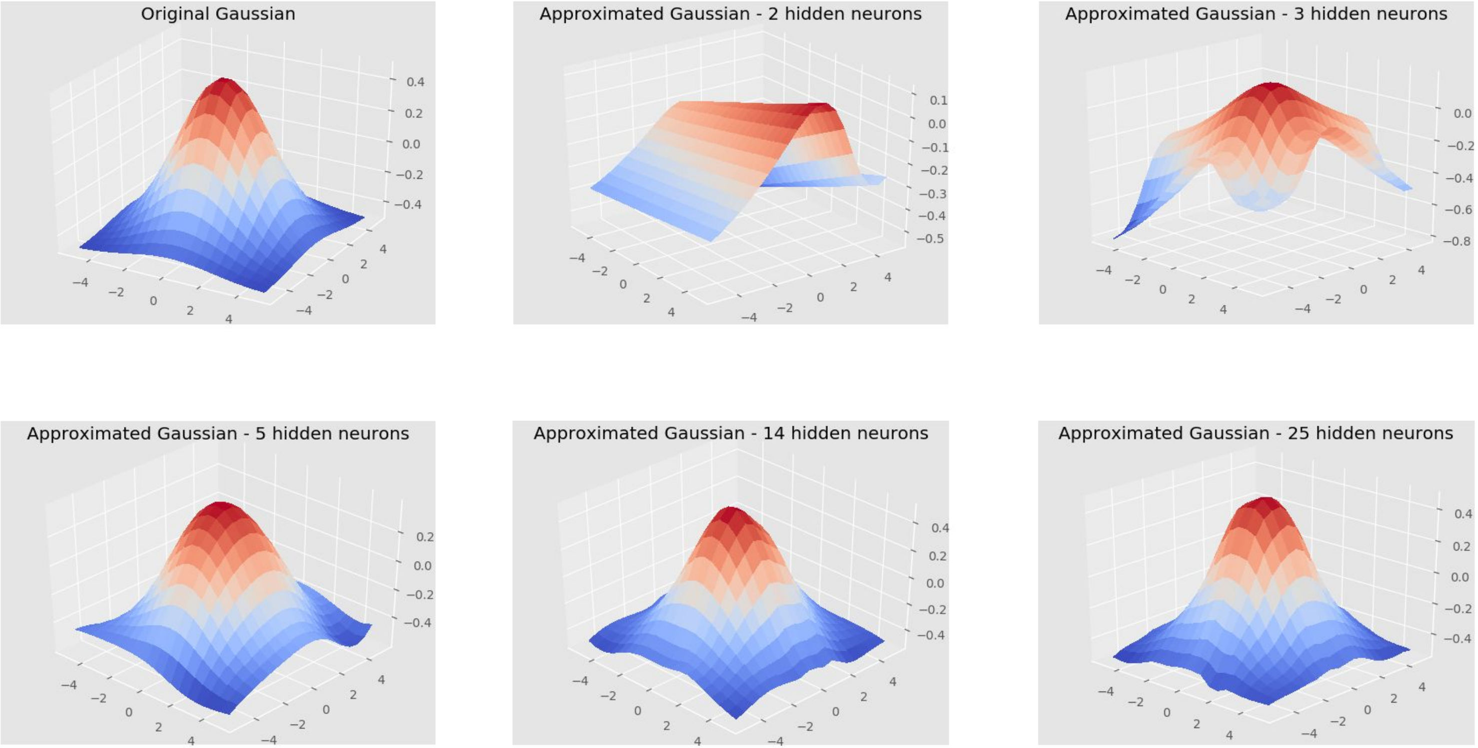
Function approximation





Left: Average MSE over 10 runs for different number of nodes in the hidden layer.
Right: Average MSE over 10 runs for different fractions of samples in training data.

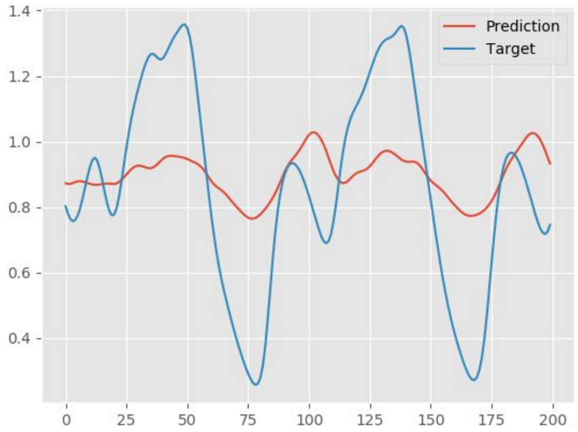
Function approximation

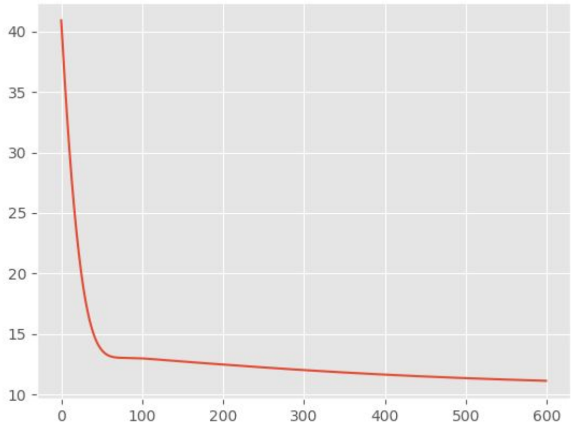


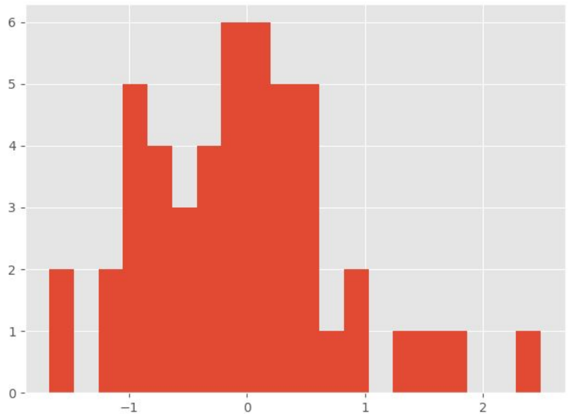
Surface plots of the approximated gaussian, with training fraction 0.7.

4
Time series prediction

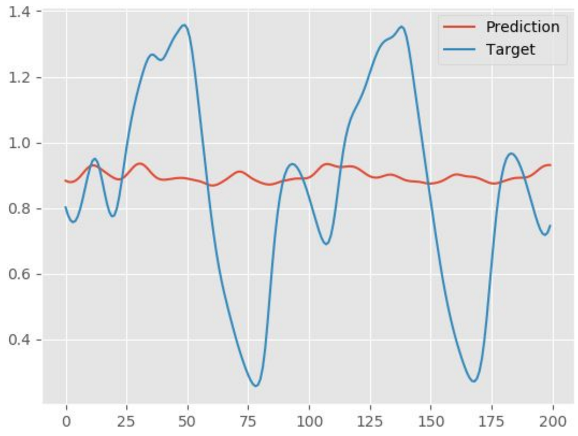
Three-layer perceptron

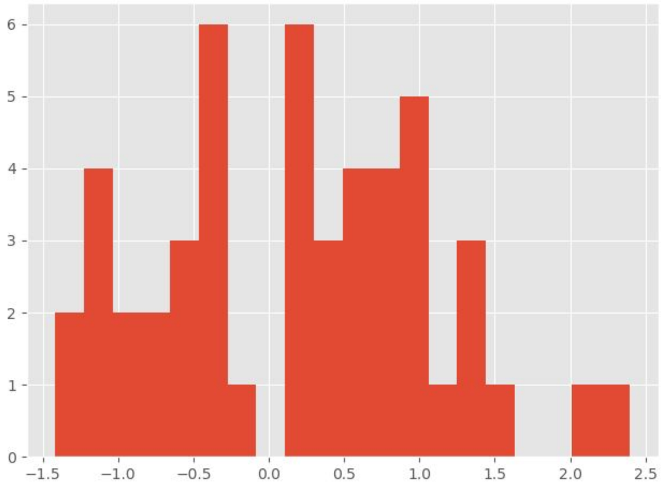


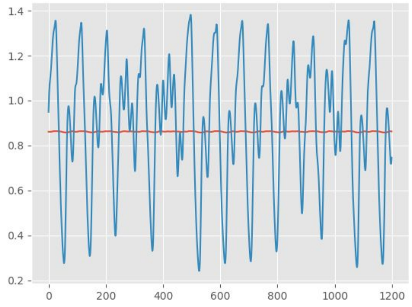




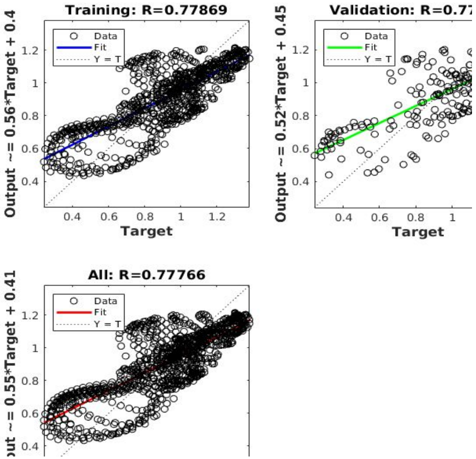
Three-layer perceptron

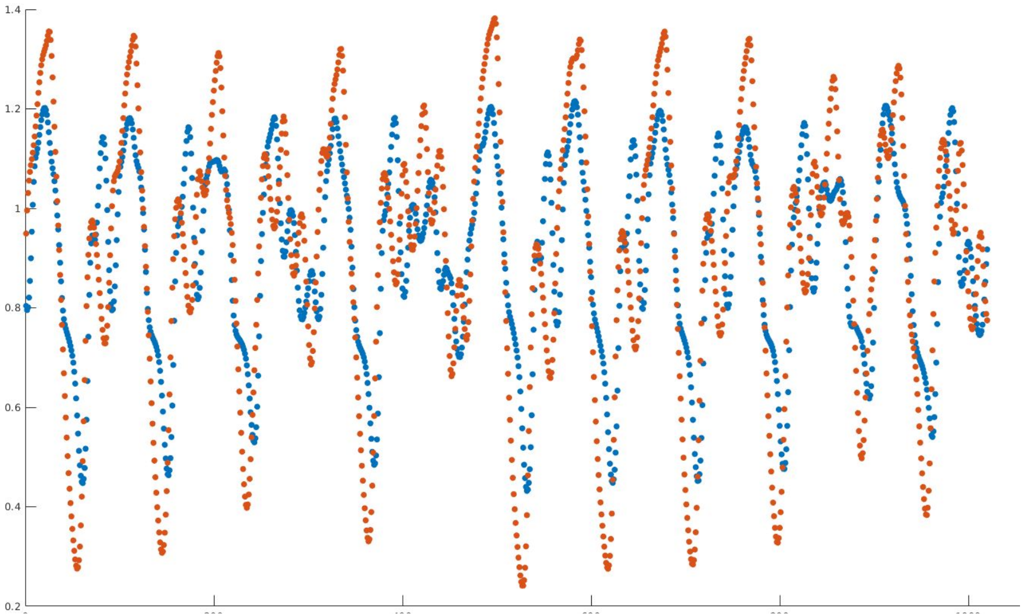




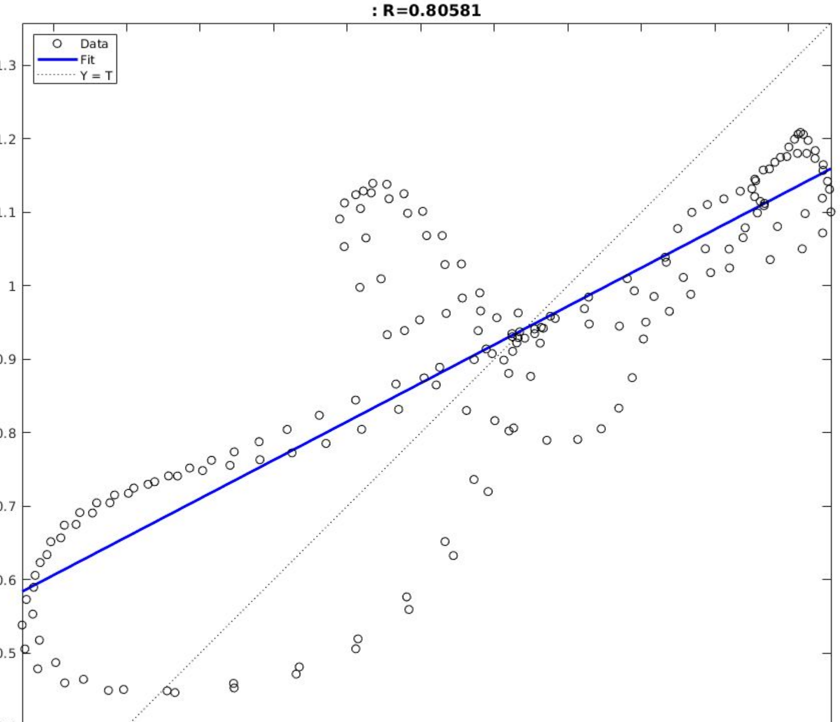


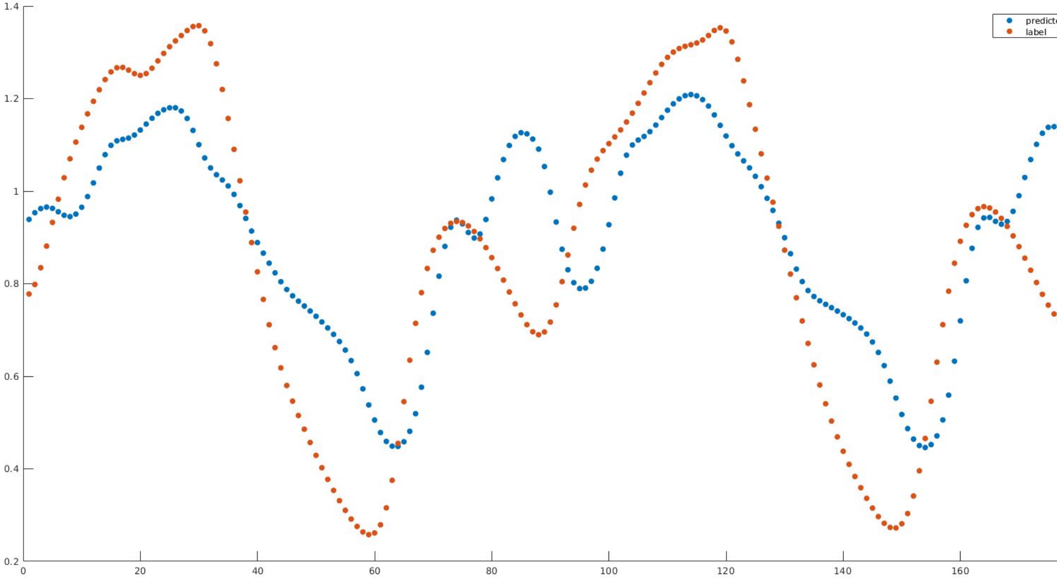
Three-layer perceptron



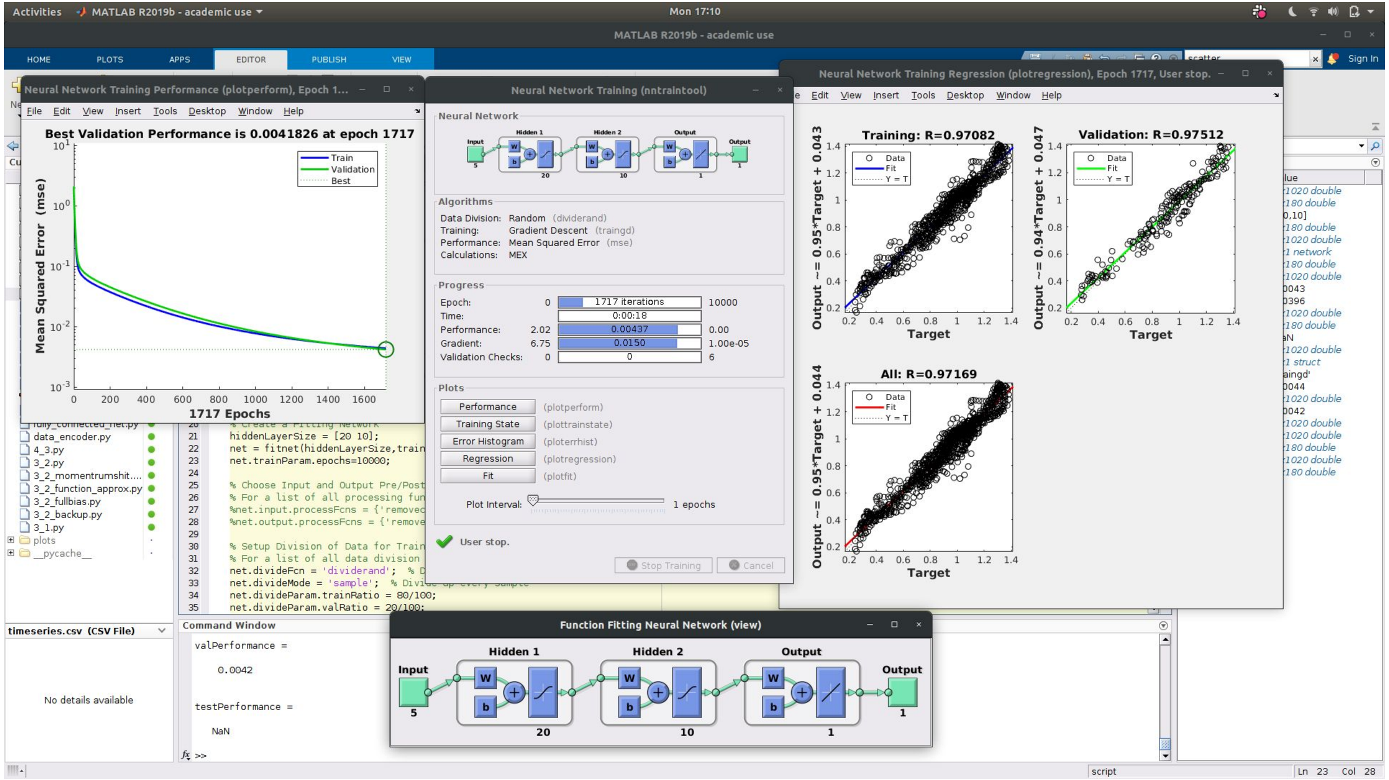


Three-layer perceptron

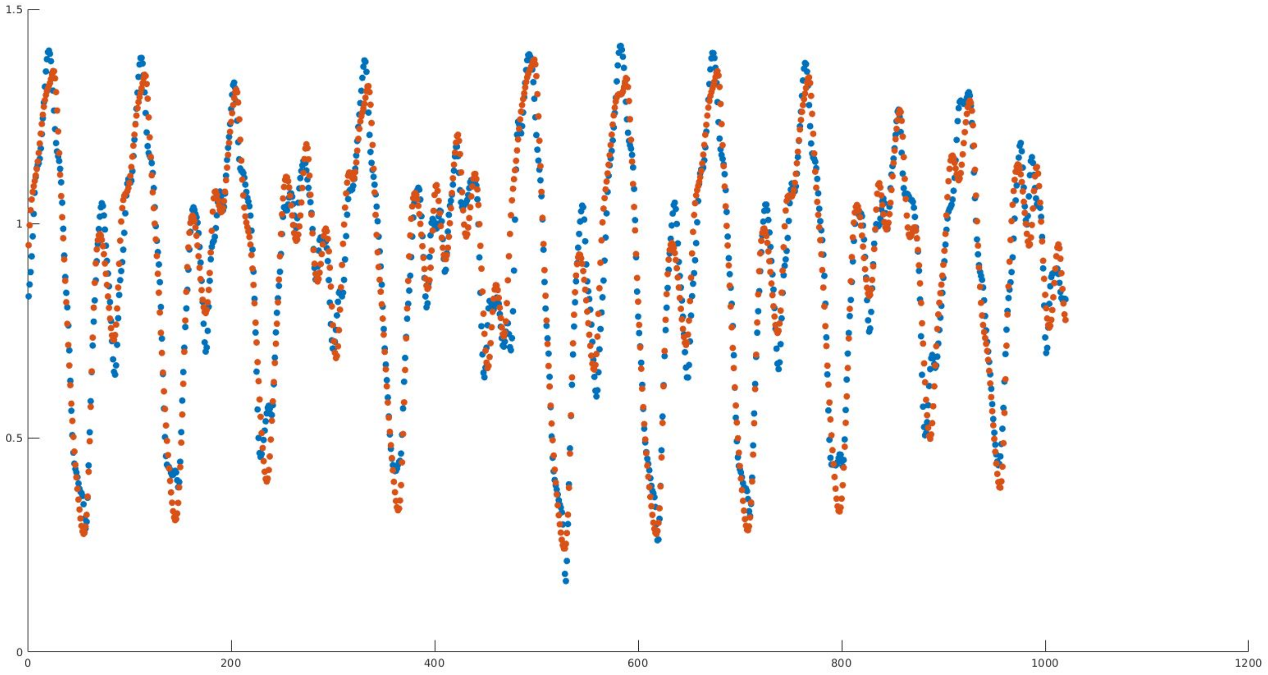


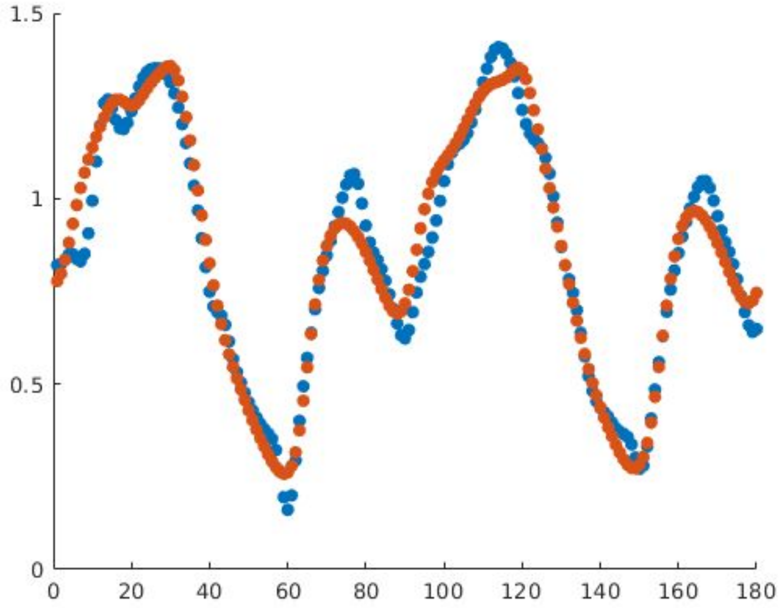


Three-layer perceptron

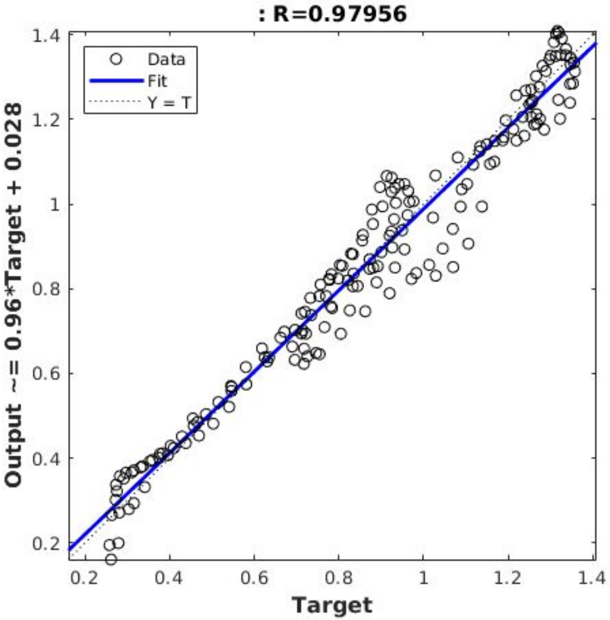


Three-layer perceptron





1.5
1
0.5
0
0
20
40
60
80
100
120
140
160
180



: R=0.97956
Data
Fit
Y = T
Output ~= 0.96*Target + 0.028
Target
0.2
0.4
0.6
0.8
1
1.2
1.4

Three-layer perceptron

Average over 100 runs for different number of nodes in the hidden layer.

Conclusions: Fastest convergence and best performance with 5 hidden nodes.